

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2003, 19:56:08 / Search time 8744 Seconds

(without alignments)
11177.161 Million cell updates/sec

Title: US-09-765-271-55

Perfect score: 2389

Sequence: 1 TTCTTACGAGTTGGACTGT.....TTAGTAGGAAAAATAAAC 2389

Scoring table: IDENTITY NUC

Gapox 10.0, Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

Genembl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_pro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hgo_hum:*
40: em_hgo_mus:*
41: em_hgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2388	100.0	2389	6	AR120265
2	2388	100.0	2389	6	BD063274
3	2388	100.0	2406	6	AX569139
4	2388	100.0	2541	1	AF291695
5	2388	100.0	8195	6	AR218862
6	2388	100.0	8195	6	BD003774
7	2388	100.0	10256	1	AE007418
8	2388	100.0	349980	6	AX571762
9	2365.6	99.0	10320	1	AE008479
10	2328.2	97.5	20035	2	SPNEU1915
11	1434.4	60.0	2535	1	AF340221
12	1374.2	57.5	2523	6	AX343072
13	1374.2	57.5	2647	6	AX343073
14	993	41.6	2457	6	AX569137
15	991	41.5	2457	1	AF318954
16	990	41.4	2517	1	AF318955
17	990	41.4	2517	6	AX568778
18	990	41.4	11931	1	AE007403
19	990	41.4	349980	6	AX571762
20	988.4	41.4	232807	2	SPNEU1901
21	987.6	41.3	2290	6	AR120270
22	987.6	41.3	2290	6	BD063279
23	980.8	41.1	2639	6	AX343074
24	962.8	40.3	12372	1	AE008464
25	869.4	36.4	2166	12	AF340222
26	835.8	35.0	75874	2	SPNEU1907
27	674	28.2	702	2	SPNEU1929
28	653.8	27.4	2359	6	AR219011
29	653.8	27.4	2359	6	BD003923
30	385.4	16.1	1342	6	AR120328
31	385.4	16.1	1342	6	BD063337
32	385.4	16.1	3117	6	AX568780
33	385.4	16.1	3120	1	AF318956
34	385.4	16.1	3120	6	AX343070
35	385.4	16.1	5048	6	AX343071
36	385.4	16.1	6867	6	AR218960
37	385.4	16.1	6867	6	BD003872
38	381.4	16.0	973	6	AR219123
39	381.4	16.0	973	6	BD004035
40	347	14.5	2528	6	AX343078
41	253.2	10.6	10312	1	AE007404
42	243.4	10.2	1684	6	AR219026
43	243.4	10.2	1684	6	BD003938
44	219.6	9.2	75248	6	AX602204
45	219.6	9.2	98050	1	SA6766854
46	214.6	9.0	2478	6	AX088372
47	214.6	9.0	10825	1	AE006623
48	213	8.9	2478	1	AB073859
49	213	8.9	10618	1	AE010110
50	211.4	8.8	1910	1	SA6290952
51	211.4	8.8	2469	6	AX088376
52	211.4	8.8	2469	6	AX607520
53	211.4	8.8	3501	1	AF062533
54	211.4	8.8	11617	6	AX602163
55	211.4	8.8	21614	1	AE014248
56	211.4	8.8	199050	1	SA6766850
57	209.8	8.8	5215	6	BD136831
58	206	8.6	21397	1	AE014279
59	205	8.6	50354	1	AE014169
60	205	8.6	323825	1	AP005146
61	201.6	8.4	1446	6	AX608407
62	188.2	7.9	492	12	AF340223
63	112.2	4.6	841	6	AR089275
64	109	4.6	841	6	AR093475
65	109	4.6	841	6	AR093475

66	102.6	4.3	1239	6	AX608408 Sequence
67	75.6	3.2	110000	2	Continuation (4 of
68	75.6	3.2	198550	2	BX004991-3
69	75.4	3.2	110000	2	PMAL13.07
70	70.6	3.0	115758	9	AC104634
71	70	2.9	250029	3	AE014839
72	69.4	2.9	110000	2	PMAL13.24
73	66.6	2.8	145265	2	AL935272
74	64	2.7	163447	9	AC104805
75	63.6	2.7	215313	9	AC100791
76	63.4	2.7	170523	9	AP002387
77	62.8	2.6	2000	6	AX65393
78	61.4	2.6	250195	3	AE014831
79	60.6	2.5	96249	9	AC116612
80	60.6	2.5	225581	2	BX537105
81	60.4	2.5	110000	2	PMAL13.06
82	60.4	2.5	253924	3	AE014822
83	60.2	2.5	51939	9	AC099341
84	60.2	2.5	213544	2	BX510939
85	60	2.5	61052	2	AC123513
86	60	2.5	136240	3	AC117070
87	59.4	2.5	250029	3	AE014820
88	58.6	2.5	272545	2	AC090533
89	58.2	2.4	252394	3	AE014833
90	58	2.4	94434	5	AL929250
91	57.8	2.4	249995	3	AE014840
92	57.8	2.4	253305	3	PMAL13.07
93	57.4	2.4	7218	6	166494
94	57.2	2.4	110000	2	PMALBP1.02
95	57.2	2.4	114226	2	AC004710
96	57.2	2.4	196490	2	AC005507
97	57.2	2.4	250421	3	AE014849
98	57	2.4	84618	5	AL591365
99	57	2.4	203138	5	BX530070
100	56.4	2.4	151343	5	AL929536

ALIGNMENTS

Qy	121	GCCTGAAGAAATCAATGCTGAGCAAAATGCTCATCAAGATAACAGACCAAGGCTATGTCAC	180
Db	121	GCCTGAAGAAATCAATGCTGAGCAAAATGCTCATCAAGATAACAGACCAAGGCTATGTCAC	180
Qy	181	TTTCATATGGGACCACTATCTATTATTACAAATGTAAGTTCTTATAGCCTATCATATG	240
Db	181	TTTCATATGGGACCACTATCTATTATTACAAATGTAAGTTCTTATAGCCTATCATATG	240
Qy	241	TGAAGAAATTAATCAATGAAGATCCAAATCAATTAAGTAAAGATGATTTGTTAATGA	300
Db	241	TGAAGAAATTAATCAATGAAGATCCAAATCAATTAAGTAAAGATGATTTGTTAATGA	300
Qy	301	GGTCAAGGATGATATGTTTATCAAGATGAGAAAATCTATGTTTACCTTAAGATGC	360
Db	301	GGTCAAGGATGATATGTTTATCAAGATGAGAAAATCTATGTTTACCTTAAGATGC	360
Qy	361	TGCCACGCGGATTAAGTCCGTACAAAAGAGAAATCAATGACAAAACCAAGGCATAG	420
Db	361	TGCCACGCGGATTAAGTCCGTACAAAAGAGAAATCAATGACAAAACCAAGGCATAG	420
Qy	421	TCAACATCGTGAAGTGAACCTCCAAAGAAACGATGCTGTTGCTTGGCAGCTTGCA	480
Db	421	TCAACATCGTGAAGTGAACCTCCAAAGAAACGATGCTGTTGCTTGGCAGCTTGCA	480
Qy	481	AGGACGCTATACATACAGATGATGTTATCTTTATGCTTCTGATATCATAGAGATAC	540
Db	481	AGGACGCTATACATACAGATGATGTTATCTTTATGCTTCTGATATCATAGAGATAC	540
Qy	541	TGCTGATGCTTATATGCTTCTCTCATGAGATCATTAACCTTACCTTAAGATGAGTT	600
Db	541	TGCTGATGCTTATATGCTTCTCTCATGAGATCATTAACCTTACCTTAAGATGAGTT	600
Qy	601	ATCAGCTAGCAGATGCTGCTGTCAGAAAGCCTTCTATGCTGTCAGAAATCTGTCAA	660
Db	601	ATCAGCTAGCAGATGCTGCTGTCAGAAAGCCTTCTATGCTGTCAGAAATCTGTCAA	660
Qy	661	TTCAAGAACCTATGCTCCGACAAATAGCATTAACCTTCAAGAAACAACTGGGATACCTTC	720
Db	661	TTCAAGAACCTATGCTCCGACAAATAGCATTAACCTTCAAGAAACAACTGGGATACCTTC	720
Qy	721	TGTAAGCAATCCAGAACTACAAATTAACACAGCAACCAACCACTTAACAGTCA	780
Db	721	TGTAAGCAATCCAGAACTACAAATTAACACAGCAACCAACCACTTAACAGTCA	780
Qy	781	AGCAAGTCAAGATGATGATGATGATGCTCTTGAACAGCTTACAAAGCTGCTTGAG	840
Db	781	AGCAAGTCAAGATGATGATGATGATGCTCTTGAACAGCTTACAAAGCTGCTTGAG	840
Qy	841	TCAACGACATGTAGATCTGATGCTTGTCTTGTATCCAGCACAAATCACAAGTGCAGC	900
Db	841	TCAACGACATGTAGATCTGATGCTTGTCTTGTATCCAGCACAAATCACAAGTGCAGC	900
Qy	901	AGCTAGAGGTTTCAAGTGCACACGAGATCATTAACCTTACCTTACTTCAAT	960
Db	901	AGCTAGAGGTTTCAAGTGCACACGAGATCATTAACCTTACCTTACTTCAAT	960
Qy	961	GTCGAATTTGGAAGAAGAAATGCTGATTAATCCCTTGATGCTTATGCTTCAACATTC	1020
Db	961	GTCGAATTTGGAAGAAGAAATGCTGATTAATCCCTTGATGCTTATGCTTCAACATTC	1020
Qy	1021	GGTACACAGATTCAGGCGACAGAACCAACAGTCCAAACCGACTCCGGAACCTAGTCCAGG	1080
Db	1021	GGTACACAGATTCAGGCGACAGAACCAACAGTCCAAACCGACTCCGGAACCTAGTCCAGG	1080
Qy	1081	CCGCAACCTGACCAATCTTAAATTAAGTCAAAATTTCTTGTGTTAGTCACTGCT	1140
Db	1081	CCGCAACCTGACCAATCTTAAATTAAGTCAAAATTTCTTGTGTTAGTCACTGCT	1140
Qy	1141	ACGAAAAGTGGGGAAGATATGATTCGAAGAAAGGCAATCTCTGTTATGCTTGTTC	1200
Db	1141	ACGAAAAGTGGGGAAGATATGATTCGAAGAAAGGCAATCTCTGTTATGCTTGTTC	1200
Qy	1201	GAAGATTTACATCTGAAATCTTAAATCTTGAAGCAAGTTATCAAAAAGAGAG	1260

RESULT 1
 ARI120265
 LOCUS ARI120265
 DEFINITION Sequence 55 from patent US 6159469.
 ACCESSION ARI120265
 VERSION ARI120265.1 GI:14103841
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 2389)
 AUTHORS Choi, G.H., Kunsch, C.A., Barash, S.C., Dillon, P.J., Dougherty, B.,
 TITLE Streptococcus pneumoniae antigens and vaccines
 JOURNAL Patent: US 6159469-A 55 12-DEC-2000;
 FEATURES
 source 1..2389
 location/Qualifiers
 BASE COUNT 830 a 461 c 486 g 611 t 1 others
 ORIGIN

Query Match 100.0%; Score 2389; DB 6; Length 2389;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTCTTACGAGTTGGACTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTCTTCTTA 60
 Db 1 TTCTTACGAGTTGGACTGTATCAAGCTAGAACGGTTAAGGAAATAATCGTGTCTTCTTA 60
 Qy 61 TATGATGGAAGAACGAGCGCAAAACGAGAGATTGCTCTGATGAGGTTAGCA 120
 Db 61 TATGATGGAAGAACGAGCGCAAAACGAGAGATTGCTCTGATGAGGTTAGCA 120

```

Db      1201  GAAAGATTACATCTGAAAGCTTTAAATCTTGAAGCAAGTATCAAAAACAAGAG 1260
Oy      1261  TGTTCACACACATTTAACTGCTAAAAAAGTGTGCTCCGAGCAAGAAATTTA 1320
Db      1261  TGTTCACACACATTTAACTGCTAAAAAAGTGTGCTCCGAGCAAGAAATTTA 1320
Oy      1321  TGATTAAGCATATATCTGTCTAACTGAGCTATAAAGCCTTGTGTAATAGGTCG 1380
Db      1321  TGATTAAGCATATATCTGTCTAACTGAGCTATAAAGCCTTGTGTAATAGGTCG 1380
Oy      1381  TAATTCGATTTTCCAAAGCCTTAGACAAATTTTGAACCGTTGAATGATGATCGA 1440
Db      1381  TAATTCGATTTTCCAAAGCCTTAGACAAATTTTGAACCGTTGAATGATGATCGA 1440
Oy      1441  TAAAGAAAATGTGATGATTTATGATTCCTAGACCAATTAACCATCCAGAGG 1500
Db      1441  TAAAGAAAATGTGATGATTTATGATTCCTAGACCAATTAACCATCCAGAGG 1500
Oy      1501  ACTTGCCAAACCAAAATTTCTCAATTTAGTATCTGAAGAAGAGTTCGATTTGCT 1560
Db      1501  ACTTGCCAAACCAAAATTTCTCAATTTAGTATCTGAAGAAGAGTTCGATTTGCT 1560
Oy      1561  AGCTGATTAAGTATCAAGCTCAGATGCTTACATTTTGAATGAACATGATATATC 1620
Db      1561  AGCTGATTAAGTATCAAGCTCAGATGCTTACATTTTGAATGAACATGATATATC 1620
Oy      1621  TGAAGAGATGATATGATTAAGCTCATATGAGGCCATGCTCATGATTTGAAAA 1680
Db      1621  TGAAGAGATGATATGATTAAGCTCATATGAGGCCATGCTCATGATTTGAAAA 1680
Oy      1681  CCTTCTGATGAAGAAAAGTTGAGCTCAGCTCAAGCTTACTTAAAGAAAAGTATC 1740
Db      1681  CCTTCTGATGAAGAAAAGTTGAGCTCAGCTCAAGCTTACTTAAAGAAAAGTATC 1740
Oy      1741  TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGAGATGTCAGCAGCTATTT 1800
Db      1741  TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGAGATGTCAGCAGCTATTT 1800
Oy      1801  CAATGCTGGAAGGGGAAAAAGAAATTCACCTCGTTCGATCCATATATGTTGAGA 1860
Db      1801  CAATGCTGGAAGGGGAAAAAGAAATTCACCTCGTTCGATCCATATATGTTGAGA 1860
Oy      1861  TACAGTTGAGTTAAAAAGCTAATTTGATTTCTCTAATGAGATCATTAACATAT 1920
Db      1861  TACAGTTGAGTTAAAAAGCTAATTTGATTTCTCTAATGAGATCATTAACATAT 1920
Oy      1921  TAAATTTGCTGTTGATGATCAACATACAAAGCTCCAAATGCGTATACCTTGAGA 1980
Db      1921  TAAATTTGCTGTTGATGATCAACATACAAAGCTCCAAATGCGTATACCTTGAGA 1980
Oy      1981  TTTTGTGAGAGATTAATTAAGTATTAATTTCTCTAATGAGATCATTAACATAT 2040
Db      1981  TTTTGTGAGAGATTAATTAAGTATTAATTTCTCTAATGAGATCATTAACATAT 2040
Oy      2041  TGGATGGGGCAATGCGCAGTGAAGATGTTAGGCAAGAAAGCAACAGAGAGATCCA 2100
Db      2041  TGGATGGGGCAATGCGCAGTGAAGATGTTAGGCAAGAAAGCAACAGAGAGATCCA 2100
Oy      2101  TAAAGACTTCAAGCGGATGAAGAGCCAGTAGAGAAACAACCTGTCAGCCAGAATCC 2160
Db      2101  TAAAGACTTCAAGCGGATGAAGAGCCAGTAGAGAAACAACCTGTCAGCCAGAATCC 2160
Oy      2161  TCAAGTAGAGATGAAAAAGTAGAAGCCCACTCAAGAAAGAGAAATTTTGTTCGGA 2220
Db      2161  TCAAGTAGAGATGAAAAAGTAGAAGCCCACTCAAGAAAGAGAAATTTTGTTCGGA 2220
Oy      2221  AGTAACGAGATTTAGTCTGAAGCCCAATGCAACAGAACTAGAGCTTTTGAATTA 2280
Db      2221  AGTAACGAGATTTAGTCTGAAGCCCAATGCAACAGAACTAGAGCTTTTGAATTA 2280
Oy      2281  TTTGACTCTTCAATATGATTAACATATGATCATGGCAGAGAGAAATTAATCTTGC 2340

```

```

Db      2281  TTTGACTCTTCAATATTAATGATTAACATATGATCATGCGACAGAGAAATTAATCTTGC 2340
Oy      2341  GTTGTAAAGAGATTAATCTTATCTGTAAGTAAAGAAATTAAC 2389
Db      2341  GTTGTAAAGAGATTAATCTTATCTGTAAGTAAAGAAATTAAC 2389

RESULT 2
BD063274
LOCUS      2389 bp      DNA      linear      PAT 27-AUG-2002
DEFINITION Streptococcus pneumoniae antigens and vaccines.
ACCESSION BD063274.1 GI:22608877
VERSION    JP 2001505415-A/28.
KEYWORDS   Streptococcus pneumoniae antigens and vaccines
SOURCE     Streptococcus pneumoniae
ORGANISM   Streptococcus pneumoniae
REFERENCE  1 (bases 1 to 2389)
AUTHORS   Kunsch, C.A., Choi, G.H., Johnson, S.L. and Hromocky, A.
TITLE     Streptococcus pneumoniae antigens and vaccines
JOURNAL    Patent: JP 2001505415-A 28 24-APR-2001;
HUMAN GENOME SCIENCES INC
COMMENT    PN JP 2001505415-A/28
PD         24-APR-2001
PF         30-OCT-1997 JP 1998520667
PR         31-OCT-1996 US 60/029960
PI         CHARLES A KUNSCH, GIL H CHOI, SYDNOR L JOHNSON, ALEX HROMOCKY
          C12N15/31, C12N5/18, C12N1/21, C07K14/315, C12Q1/68, A61K39/09, PC
          G01N33/569,
          PC G01N33/68
          CC Strandedness: Double;
          CC Topology: Linear;
          FH Key Location/Qualifiers.

FEATURES             Location/Qualifiers
     source            1..2389
                        /organism="Streptococcus pneumoniae"
                        /mol_type="genomic DNA"
                        /db_xref="taxon:32644"
BASE COUNT    830 a 461 c 486 g 611 t 1 others
ORIGIN
Query Match      100.0%; Score 2389; DB 6; Length 2389;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 TTCTTACGAGTTGGAGCTGATCAAGCTTAGAAGCGTTAAGAAATTAATCGTTTCTTA 60
Db      1 TTCTTACGAGTTGGAGCTGATCAAGCTTAGAAGCGTTAAGAAATTAATCGTTTCTTA 60
Oy      61 TATAGATGAAAGCAAGGAGCGCAAAAAACGAGAAATTTGACTCCTGATGAGGTTAGCA 120
Db      61 TATAGATGAAAGCAAGGAGCGCAAAAAACGAGAAATTTGACTCCTGATGAGGTTAGCA 120
Oy      121 GCGTGAAGAAATCAATGCTGAGCAAAATGTCATCAAGATTAACAGACCAAGGCTATGTCAC 180
Db      121 GCGTGAAGAAATCAATGCTGAGCAAAATGTCATCAAGATTAACAGACCAAGGCTATGTCAC 180
Oy      181 TGCATGAGCGACACTATCATTTATTACATGTAAGGTTCTTATGACGCTATCATGAG 240
Db      181 TGCATGAGCGACACTATCATTTATTACATGTAAGGTTCTTATGACGCTATCATGAG 240
Oy      241 TGAAGATTAATCAAGAAAGATCCAACTATAGCTTAAAGATGAGATATTGTTAATGA 300
Db      241 TGAAGATTAATCAAGAAAGATCCAACTATAGCTTAAAGATGAGATATTGTTAATGA 300
Oy      301 GGTCAAGGATGATATGTTATCAAGTATGAGTGAAGAAATTAATGTTTAAAGATGC 360
Db      301 GGTCAAGGATGATATGTTATCAAGTATGAGTGAAGAAATTAATGTTTAAAGATGC 360
Oy      361 TGCCACGCGGATTAACGTCGTTCAAAAAGAGAAATCAATGACCAAAAACAAAGACATAG 420
Db      361 TGCCACGCGGATTAACGTCGTTCAAAAAGAGAAATCAATGACCAAAAACAAAGACATAG 420

```

QY 421 TCACATCGTGAAGGTGGAACCTCCAGAAACGATGCTGCTGCTGCTGCGACGTTGCCA 480
 Db 421 TCACATCGTGAAGGTGGAACCTCCAGAAACGATGCTGCTGCTGCTGCGACGTTGCCA 480
 QY 481 AGGAGCGTATCTACAGATGATGCTTATATCTTTAATGCTTCTGATATCAGAGATAC 540
 Db 481 AGGAGCGTATCTACAGATGATGCTTATATCTTTAATGCTTCTGATATCAGAGATAC 540
 QY 541 TGGTATGCTTATATGCTTCTCATGAGATCATTAACATTCCTTAAGATGATG 600
 Db 541 TGGTATGCTTATATGCTTCTCATGAGATCATTAACATTCCTTAAGATGATG 600
 QY 601 ATCACTAGCGAGTGGCTGTCAGAAAGCTTCTCATGCTGAGGAAATCTGTCAA 660
 Db 601 ATCACTAGCGAGTGGCTGTCAGAAAGCTTCTCATGCTGAGGAAATCTGTCAA 660
 QY 661 TTCAGAACTTATCGCCGACAAATAGCGATACCTTCAGAAACAACTGGGTACCTTC 720
 Db 661 TTCAGAACTTATCGCCGACAAATAGCGATACCTTCAGAAACAACTGGGTACCTTC 720
 QY 721 TGTATGCAATCCAGGAACTACAAATATCTAACAGCAACAGCAACAGCAACAGTCA 780
 Db 721 TGTATGCAATCCAGGAACTACAAATATCTAACAGCAACAGCAACAGCAACAGTCA 780
 QY 781 AGCAAGTCAAGTATGACATTTGATATGCTCTTGAACAGCTCTCAAACTGCCCTTGA 840
 Db 781 AGCAAGTCAAGTATGACATTTGATATGCTCTTGAACAGCTCTCAAACTGCCCTTGA 840
 QY 841 TCACGACATGATGATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
 Db 841 TCACGACATGATGATCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
 QY 901 AGCTAGAGTGTGAGTGGCCACAGAGATCATTAACATTCCTTCACTCTCAAT 960
 Db 901 AGCTAGAGTGTGAGTGGCCACAGAGATCATTAACATTCCTTCACTCTCAAT 960
 QY 961 GTCTGAATTTGAGAGACGAATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 Db 961 GTCTGAATTTGAGAGACGAATCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
 QY 1021 GGTACCGATTCAGAGCCAGAAACCAAGTCCACACGACTCCGGAACCTAGTCCAG 1080
 Db 1021 GGTACCGATTCAGAGCCAGAAACCAAGTCCACACGACTCCGGAACCTAGTCCAG 1080
 QY 1081 CCGGCAACCTGACCAATCTTAAATAGCTCAATCTTCTGCTGCTGCTGCTGCTGCT 1140
 Db 1081 CCGGCAACCTGACCAATCTTAAATAGCTCAATCTTCTGCTGCTGCTGCTGCTGCT 1140
 QY 1141 ACGAAAGTTGGGAGAGATATGATTCGAAGAAAGGCACTCTCTGCTGCTGCTGCTG 1200
 Db 1141 ACGAAAGTTGGGAGAGATATGATTCGAAGAAAGGCACTCTCTGCTGCTGCTGCTG 1200
 QY 1201 GAAAGATTTACATCTGAACTGTTAAATCTTGAAGCAAGTTATCAAAACAGAGAG 1260
 Db 1201 GAAAGATTTACATCTGAACTGTTAAATCTTGAAGCAAGTTATCAAAACAGAGAG 1260
 QY 1261 TGTTCACACACTTAACTGCTAAAGAAAGAAATGCTGCTGCTGCTGCTGCTGCTGCT 1320
 Db 1261 TGTTCACACACTTAACTGCTAAAGAAAGAAATGCTGCTGCTGCTGCTGCTGCTGCT 1320
 QY 1321 TGATTAAGCATATATCTGTTAACTGAGCTCATTAAGCTGTTGTTGTTGTTGTTG 1380
 Db 1321 TGATTAAGCATATATCTGTTAACTGAGCTCATTAAGCTGTTGTTGTTGTTGTTG 1380
 QY 1381 TAAATTCGATTTCCAGCTTAGACAAATATATAGAACGCTTGAATGATGATGATG 1440
 Db 1381 TAAATTCGATTTCCAGCTTAGACAAATATATAGAACGCTTGAATGATGATGATG 1440
 QY 1441 TAAAGAAATTTGATGATGATTTATTTGCAATCTTACACCAATTAACCAACAGAG 1500
 Db 1441 TAAAGAAATTTGATGATGATTTATTTGCAATCTTACACCAATTAACCAACAGAG 1500
 QY 1501 ACTTGCAAAACCAATCTCAAAATGATATATGAAAGCAAGTTGCTATGCTCAAT 1560

Db 1501 ACTTGCAAAACCAATCTCAAAATGATATGAAAGCAAGTTGCTATGCTCAAT 1560
 QY 1561 AGCTGATGATATCAACGCTCAGATGCTTACATTTTGAAGCAATGATATTAACAGTGA 1620
 Db 1561 AGCTGATGATATCAACGCTCAGATGCTTACATTTTGAAGCAATGATATTAACAGTGA 1620
 QY 1621 TGAAGAGATGATATGATGCTCAGATGCTTACATTTTGAAGCAATGATATTAACAGTGA 1680
 Db 1621 TGAAGAGATGATATGATGCTCAGATGCTTACATTTTGAAGCAATGATATTAACAGTGA 1680
 QY 1681 CTTTCTGATGAGAAAGTTGCACTCAAGCTTATATCTTAAGAAAGGATATCTTACC 1740
 Db 1681 CTTTCTGATGAGAAAGTTGCACTCAAGCTTATATCTTAAGAAAGGATATCTTACC 1740
 QY 1741 TCCATCTCCAGACCAATGTTAAGAAATCCAACTGAGATGATGCTGAGCTATTTA 1800
 Db 1741 TCCATCTCCAGACCAATGTTAAGAAATCCAACTGAGATGATGCTGAGCTATTTA 1800
 QY 1801 CAATCTGATGAGAAAGGGAAGAAATTCATCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
 Db 1801 CAATCTGATGAGAAAGGGAAGAAATTCATCTGCTGCTGCTGCTGCTGCTGCTGCTG 1860
 QY 1861 TACAGTGAAGTTAAAAAGGTAATTTGATATTTCTCATTAAGATCATTAACATTAAT 1920
 Db 1861 TACAGTGAAGTTAAAAAGGTAATTTGATATTTCTCATTAAGATCATTAACATTAAT 1920
 QY 1921 TAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
 Db 1921 TAAATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1980
 QY 1981 TTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2040
 Db 1981 TTTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2040
 QY 2041 TGGATGGGCAATCCAGTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2100
 Db 2041 TGGATGGGCAATCCAGTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2100
 QY 2101 TAAAGCTTCAAAACCGATGAGAGGCAAGTGAAGAAACCTGCTGAGCAGAAAGTCCC 2160
 Db 2101 TAAAGCTTCAAAACCGATGAGAGGCAAGTGAAGAAACCTGCTGAGCAGAAAGTCCC 2160
 QY 2161 TCAAGTGAAGCTGAGAAAGTGAAGCCCACTCAAGAGCAAGATTTTCTTGGCAA 2220
 Db 2161 TCAAGTGAAGCTGAGAAAGTGAAGCCCACTCAAGAGCAAGATTTTCTTGGCAA 2220
 QY 2221 AGTAAGCGATTTCTGATGAGAGCCAAATGCAACGAAATCTTACCTGTTACCAATTA 2280
 Db 2221 AGTAAGCGATTTCTGATGAGAGCCAAATGCAACGAAATCTTACCTGTTACCAATTA 2280
 QY 2281 TTTGACCTTCAATTTATGATTAACAATGATATGATGATGATGATGATGATGATGATG 2340
 Db 2281 TTTGACCTTCAATTTATGATTAACAATGATATGATGATGATGATGATGATGATGATG 2340
 QY 2341 GTTGTAAAGAAATATCTTCAATCTGTAAGTAAGAAAGAAATTAAC 2389
 Db 2341 GTTGTAAAGAAATATCTTCAATCTGTAAGTAAGAAAGAAATTAAC 2389

RESULT 3
 AX569139 2406 bp DNA linear PAT 29-NOV-2002
 LOCUS Sequence 2347 from Patent WO02077021.
 DEFINITION AX569139
 ACCESSION AX569139.1 GI:26002636
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Streptococcus pneumoniae
 Streptococcus pneumoniae
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 REFERENCE
 1 Masignani, V., Tettelin, H. and Fraser, C.

TITLE Streptococcus pneumoniae proteins and nucleic acids
JOURNAL Patent: WO 02077021-A 2347 03-OCT-2002;
Chiron Spa (IT) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES Location/Qualifiers
source 1..2406
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/db_xref="taxon:1313"
BASE COUNT 834 a 461 c 490 g 621 t
ORIGIN
Query Match 100.0%; Score 2388; DB 6; Length 2406;
Beet Local Similarity 100.0%; Pred. No. 0;
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TTCTTACGAGTGGGAGCTGATCAAGCTGAACGGTTAGGAAATTAATCGGTTCCCA 60
DB 18 TTCTTACGAGTGGGAGCTGATCAAGCTGAACGGTTAGGAAATTAATCGGTTCCCA 77
QY 61 TATAGATGGAACCAAGCGCAACCAAAACGAGAAATTTGACTCTGATGAGTTAGCA 120
DB 78 TATAGATGGAACCAAGCGCAACCAAAACGAGAAATTTGACTCTGATGAGTTAGCA 137
QY 121 GCGTGAAGAAATCAATGCTGAGCAAAATCGTCATCAAGATACAGACCAAGCTATGTCAC 180
DB 138 GCGTGAAGAAATCAATGCTGAGCAAAATCGTCATCAAGATACAGACCAAGCTATGTCAC 197
QY 181 TTCACATGGCGACCACTATCATTTATTAAGATAGTAAAGTTCTTTAGAGCTATCATCG 240
DB 198 TTCACATGGCGACCACTATCATTTATTAAGATAGTAAAGTTCTTTAGAGCTATCATCG 257
QY 241 TGAAGAAATTTACTCATGAAGATCCAAACTATTAAGCTAAAGATGAGATATTGTTAATGA 300
DB 258 TGAAGAAATTTACTCATGAAGATCCAAACTATTAAGCTAAAGATGAGATATTGTTAATGA 317
QY 301 GGTCAAGGTGATATGTTATCAAGTAGATGAAATTAATGTTTACCTTAAGAGATGC 360
DB 318 GGTCAAGGTGATATGTTATCAAGTAGATGAAATTAATGTTTACCTTAAGAGATGC 377
QY 361 TGCCACGCGGATACGTCCTGACAAAGAAAGAAATCATTCGACAAACCAAGACATAG 420
DB 378 TGCCACGCGGATACGTCCTGACAAAGAAAGAAATCATTCGACAAACCAAGACATAG 437
QY 421 TCACATCGTGAAGTGAATCCCAAGAAAGATGGTGTGCTTGCTGGACGTTCCGA 480
DB 438 TCACATCGTGAAGTGAATCCCAAGAAAGATGGTGTGCTTGCTGGACGTTCCGA 497
QY 481 AGGACGCTATACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
DB 498 AGGACGCTATACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 557
QY 541 TGGTATGCTTATATCGTTCCTCATGAGATCATTAACCTTACCTTCAAGATGAGTT 600
DB 558 TGGTATGCTTATATCGTTCCTCATGAGATCATTAACCTTACCTTCAAGATGAGTT 617
QY 601 ATCAGCTAGCGAGTTGGCTGTGACGAAGCCTTCCATCTGGTCGAGGAATCTGCAAA 660
DB 618 ATCAGCTAGCGAGTTGGCTGTGACGAAGCCTTCCATCTGGTCGAGGAATCTGCAAA 677
QY 661 TTCAAGAACTTATCGCCGCAAAATATAGCATTAACCTTCAAGAACTTGGGATACCTTC 720
DB 678 TTCAAGAACTTATCGCCGCAAAATATAGCATTAACCTTCAAGAACTTGGGATACCTTC 737
QY 721 TGTATAGCAATCCAGAACTTCAAAATTAATAACAAGCAACACAGACATTAACATGA 780
DB 738 TGTATAGCAATCCAGAACTTCAAAATTAATAACAAGCAACACAGACATTAACATGA 797
QY 781 AGCAAGTCAAAATTAATGACATTAATGATGCTCTTGAAGACGCTCTCAAACTGCTTTGAG 840
DB 798 AGCAAGTCAAAATTAATGACATTAATGATGCTCTTGAAGACGCTCTCAAACTGCTTTGAG 857
QY 841 TCAACGACATGTAGATCTGATGCGCTTGTCTTTGATCCAGACCAATCAAGTCGAAAC 900

DB 858 TCACGACATGTAGATCTGATGCGCTTGTCTTTGATCCAGACCAAAATCACAAAGTCGAAC 917
QY 901 AGCTAGAGTGTGAGTGCAGTGCACACGAGATCATTAACCACTTCAATCCCTTACTCAAT 960
DB 918 AGCTAGAGTGTGAGTGCAGTGCACACGAGATCATTAACCACTTCAATCCCTTACTCAAT 977
QY 961 GTCTGAATTTGGAAGAACCAATCGCTCGTATTATTCCTTCGTTATCGTTCAACCATTTG 1020
DB 978 GTCTGAATTTGGAAGAACCAATCGCTCGTATTATTCCTTCGTTATCGTTCAACCATTTG 1037
QY 1021 GGTATCAATTTCAAGCCGACAAACCAACCAAGTCCACACCGACTCCGGAACCTATGTCAGG 1080
DB 1038 GGTATCAATTTCAAGCCGACAAACCAACCAAGTCCACACCGACTCCGGAACCTATGTCAGG 1097
QY 1081 CCCGCAACCTGACCAATCTTAAATTAAGCTCAAAATCTTCTTTGTTAGCAGCTGGT 1140
DB 1098 CCCGCAACCTGACCAATCTTAAATTAAGCTCAAAATCTTCTTTGTTAGCAGCTGGT 1157
QY 1141 AGGAAAGTTGGGAGAGATATGTAATTCGAAGAAAGGAGCATCTCTGTTATGTCCTTTC 1200
DB 1158 AGGAAAGTTGGGAGAGATATGTAATTCGAAGAAAGGAGCATCTCTGTTATGTCCTTTC 1217
QY 1201 GAAAGATTTTACCATCTGAAAATCTTTAAATTTCTGAAGCAAGTTATCAAAACAAAGAG 1260
DB 1218 GAAAGATTTTACCATCTGAAAATCTTTAAATTTCTGAAGCAAGTTATCAAAACAAAGAG 1277
QY 1261 TGTTCACACACTTAACTGCTTAAAGAAAGAAATGTCCTCTGTCGCAAGAAATTTTA 1320
DB 1278 TGTTCACACACTTAACTGCTTAAAGAAAGAAATGTCCTCTGTCGCAAGAAATTTTA 1337
QY 1321 TGATTAACCATTAATCTGTTAACTGAGCTCATTAAGCTTGTGTTGAAATTAAGGCTGC 1380
DB 1338 TGATTAACCATTAATCTGTTAACTGAGCTCATTAAGCTTGTGTTGAAATTAAGGCTGC 1397
QY 1381 TAATTCGATTTTCCAAAGCTTGAACAAATTAATGAACGCTTGAATGATGATGACTTA 1440
DB 1398 TAATTCGATTTTCCAAAGCTTGAACAAATTAATGAACGCTTGAATGATGATGACTTA 1457
QY 1441 TAAAGAAATTTGGATGATGATTAATTTGGCAATCTCTACCAACCAATTAACCAACGAG 1500
DB 1458 TAAAGAAATTTGGATGATGATTAATTTGGCAATCTCTACCAACCAATTAACCAACGAG 1517
QY 1501 ACTTGCAAAACCAATTTCTCAAAATTTGATTAATGAACGAGATTCCTATTGCTCAATT 1560
DB 1518 ACTTGCAAAACCAATTTCTCAAAATTTGATTAATGAACGAGATTCCTATTGCTCAATT 1577
QY 1561 AGCTGATTAATTAACCAAGTGAATGTTTACATTTTATGATGAACATGATTAATCAGTGA 1620
DB 1578 AGCTGATTAATTAACCAAGTGAATGTTTACATTTTATGATGAACATGATTAATCAGTGA 1637
QY 1621 TGAAGGATATGATTAATGATGAGCTCATTAAGGCGCATAGTCACTGATTTGGAAAAAGATAG 1680
DB 1638 TGAAGGATATGATTAATGATGAGCTCATTAAGGCGCATAGTCACTGATTTGGAAAAAGATAG 1697
QY 1681 CTTTCTGATTAAGGAAATTTGAGCTCAAGCTCATTAAGGAAATTTGATCTTCAATC 1740
DB 1698 CTTTCTGATTAAGGAAATTTGAGCTCAAGCTCATTAAGGAAATTTGATCTTCAATC 1757
QY 1741 TCCATCTCCAGACGAGATGTTAAAGCAATTCAACTGAGATTAATGACGACGATTTTA 1800
DB 1758 TCCATCTCCAGACGAGATGTTAAAGCAATTCAACTGAGATTAATGACGACGATTTTA 1817
QY 1801 CAATGCTGTAAGGAGGAAAAACCAATTCCTGCTGCTTCAATTAATGCTTGAACA 1860
DB 1818 CAATGCTGTAAGGAGGAAAAACCAATTCCTGCTGCTTCAATTAATGCTTGAACA 1877
QY 1861 TACAGTGAAGTAAAGGCTTAATTTGATTAATTTCTCATTAAGATCAATTAATTAAT 1920
DB 1878 TACAGTGAAGTAAAGGCTTAATTTGATTAATTTCTCATTAAGATCAATTAATTAAT 1937
QY 1921 TAAATTTCTTGTGTTGATGATCAACATTAACCAATGCTCAATGCTTATACCTTGAAGA 1980
DB 1938 TAAATTTCTTGTGTTGATGATCAACATTAACCAATGCTCAATGCTTATACCTTGAAGA 1997

QY 1981 TTGTTTGGACGATTAAGTACTAGCTAGAAACACCTGACGAACTCCACATTCTAATGA 2040
 Db 1998 TTTGTTTGGACGATTAAGTACTAGCTAGAAACACCTGACGAACTCCACATTCTAATGA 2057
 QY 2041 TGGATGGGGCAATGCGACGTAGCTGTGTTAGCAAGAAAGACACAGTGAAGATCCAA 2100
 Db 2058 TGGATGGGGCAATGCGACGTAGCTGTGTTAGCAAGAAAGACACAGTGAAGATCCAA 2117
 QY 2101 TTAAGAACTTCAAAAGCGGATGAAGAGCCAGTGAAGAAACCTGCTGAGCCAGAAAGTCCC 2160
 Db 2118 TAAAGAACTTCAAAAGCGGATGAAGAGCCAGTGAAGAAACCTGCTGAGCCAGAAAGTCCC 2177
 QY 2161 TCAAGTAGAGACTGAAAAAGTGAAGAGCCCACTCAAGAGAGAGAGTTTGGCTTGGCGAA 2220
 Db 2178 TCAAGTAGAGACTGAAAAAGTGAAGAGCCCACTCAAGAGAGAGAGTTTGGCTTGGCGAA 2237
 QY 2221 AGTAACGAGATTTAGCTGTAAGCCCAATGCAACAGAACTCTAGCTGTTTACGAAATGA 2280
 Db 2238 AGTAACGAGATTTAGCTGTAAGCCCAATGCAACAGAACTCTAGCTGTTTACGAAATGA 2297
 QY 2281 TTTGACTCTTCAATTTATGAGTAAACAATGATGATCATGGCAGAGAGCAAGAAATTTCTTGC 2340
 Db 2298 TTTGACTCTTCAATTTATGAGTAAACAATGATGATCATGGCAGAGAGCAAGAAATTTCTTGC 2357
 QY 2341 GTTGTAAAGGAGTAATCTTCAATCTGTAGTAGAGAAAAAATAAAC 2389
 Db 2358 GTTGTAAAGGAGTAATCTTCAATCTGTAGTAGAGAAAAAATAAAC 2406

RESULT 4
 LOCUS AF291695 2541 bp DNA linear BCT 15-MAR-2001
 DEFINITION Streptococcus pneumoniae pneumococcal histidine triad A protein
 ACCESSION AF291695
 VERSION AF291695.1 GI:13345012
 KEYWORDS
 SOURCE
 ORGANISM Streptococcus pneumoniae
 Streptococcus pneumoniae
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 REFERENCE 1 (bases 1 to 2541)
 Wisemann, T.M., Heinrichs, J.H., Adamou, J.E., Erwin, A.L., Kunisch, C.,
 Choi, G.H., Barash, S.C., Rosen, C.A., Masure, H.R., Tuomanen, E.,
 Gayle, A., Brewah, Y.A., Walsh, W., Barron, P., Lachigra, R., Hanson, M.,
 Langerman, S., Johnson, S. and Koenig, S.
 Use of a whole genome approach to identify vaccine molecules
 affording protection against Streptococcus pneumoniae infection
 JOURNAL Infect. Immun. 69 (3), 1593-1598 (2001)
 MEDLINE 2116976
 PUBMED 11179332
 TITLE 2 (bases 1 to 2541)
 REFERENCE Choi, G.H.
 Direct Submission
 Submitted (01-AUG-2000) Molecular Biology, Human Genome Sciences,
 Inc., 9410 Key West Ave., Rockville, MD 20850, USA
 JOURNAL
 FEATURES
 source
 1..2541
 /organism="Streptococcus pneumoniae"
 /mol_type="genomic DNA"
 /strain="NA"
 /db_xref="taxon:1313"
 91..2541
 /gene="phtA"
 91..2541
 /note="phtA"
 /codon_start=1
 /transl_table=11
 /product="pneumococcal histidine triad A protein"
 /protein_id="AAK19155.1"
 /db_xref="GI:13345013"

/translation="MKINKKYLVGSAALILVCSYELGLYQARTVKNRVSYIDK
 QATCKENLTPDESKREGINAEIVIKITDOGVTSRGGHYHYNKPDAIISEE
 LMKDPRYKLDIEDIVNEVKGIVYVYLKRAAHADNRTEELINROKQES
 QHRGGTFRNDGAAALARSOGRTYTDGCIIPNASDIIEDGDADIVYHGHGHIIPKX
 ELSASLEUAAEAFUSGKGLNSRTRYRQSDNSTRNWPSVNSPGITNTNSNS
 TNSQASQNDIDSLKQLYKLPLSRVIEDVLDFDQITSRARVAAPVHGSHYF
 IPIYQMSLEPERIARIILPLRVSNHWVPSDRPEQSPQPPPEPSPQAPNLIDSN
 SLVYSQLRKVEGYVEEKGISRIVYAKQLPSEYVKNLSKSKQESVSHTPAKKE
 NVAPRDEPYDKAYNLLTEAHKALFPXKKGNSDPAQDLKLERLNDSTNREKLVYL
 LAFAPATTPHERLQKPSQIETEDRYRIQADKTYTSQYITDEHDIIISDEDAVY
 TPNHGSHWIGKQSLSDKERVAAQAYTKENGILPPSPDADVKAPITDSDAAITNRYK
 GEKRIPLVRLPYVWEHTVEVKNGLIIPKDHYNINIFAMFDHTYAPNGYLTEDYF
 ATIKYVEHDERPHSNDGWNASEHLGKDKHSDENKMFKADEEVEEYEPAPREVP
 OVETEKYEQLKEAEVLVLAQVTSLSKANTETLAGRNLLIQIMDNNSIMAEKEL
 LALKSGNSPSSVSEKXIN"
 BASE COUNT 888 a 476 c 516 g 660 t 1 others
 ORIGIN
 Query Match 100.0%; Score 2388; DB 1; Length 2541;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAACGGTTAAGAAATAATCGTGTCTTA 60
 Db 150 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAACGGTTAAGAAATAATCGTGTCTTA 209
 QY 61 TATAGATGAAAAACACGCGACGCAAAAAAGGAGAAATTTGACTCTGATAGGTTACGA 120
 Db 210 TATAGATGAAAAACACGCGACGCAAAAAAGGAGAAATTTGACTCTGATAGGTTACGA 269
 QY 121 GCGTGAAGAAATCAATGCTGACCAATTCGTATCAAGATGAACAGCAAGGCTATGTCAC 180
 Db 270 GCGTGAAGAAATCAATGCTGACCAATTCGTATCAAGATGAACAGCAAGGCTATGTCAC 329
 QY 181 TTACATGCGCAGCACTATCATTTTACAATGTAAGTCTTATGACCGTATCATCAG 240
 Db 330 TTACATGCGCAGCACTATCATTTTACAATGTAAGTCTTATGACCGTATCATCAG 389
 QY 241 TGAAGATTTACTCATGAAAGTCCAACTTAAGCTAAAGATAGAGATATTGTAATGA 300
 Db 390 TGAAGATTTACTCATGAAAGTCCAACTTAAGCTAAAGATAGAGATATTGTAATGA 449
 QY 301 GGTCAAGGTGAGATATGTTATCAAGTAGATGAAATACTATGTTTACCTTAAGATGC 360
 Db 450 GGTCAAGGTGAGATATGTTATCAAGTAGATGAAATACTATGTTTACCTTAAGATGC 509
 QY 361 TGCCACGCGGATVACGTCGTACAAAGAGAAATCAATGACAAACAGAGCATAG 420
 Db 510 TGCCACGCGGATVACGTCGTACAAAGAGAAATCAATGACAAACAGAGCATAG 569
 QY 421 TCAACATCGTGAAGGTGAATCCAGAAACGATGTCGTCCTTGCGACGTTGCGCA 480
 Db 570 TCAACATCGTGAAGGTGAATCCAGAAACGATGTCGTCCTTGCGACGTTGCGCA 629
 QY 481 AGAAGCTATATCTACAGATGATGTTTATCTTAATGCTTCTGATATCATAGAGATAC 540
 Db 630 AGAAGCTATATCTACAGATGATGTTTATCTTAATGCTTCTGATATCATAGAGATAC 689
 QY 541 TGGTATGCTTATATGCTTCTCTCATGAGATCATTAACATTAATCTTAAGATGAGTT 600
 Db 690 TGGTATGCTTATATGCTTCTCTCATGAGATCATTAACATTAATCTTAAGATGAGTT 749
 QY 601 ATCAGCTAGCGAGTTGGCTGTGACAGAAAGCTTCTTATCTGCTGAGAGAAATCTGTCAA 660
 Db 750 ATCAGCTAGCGAGTTGGCTGTGACAGAAAGCTTCTTATCTGCTGAGAGAAATCTGTCAA 809
 QY 661 TTCAAGAACCTATGCGCGACAAATATAGATTAACCTTCAAGAACTGAGTACCTTC 720
 Db 810 TTCAAGAACCTATGCGCGACAAATATAGATTAACCTTCAAGAACTGAGTACCTTC 869
 QY 721 TGTAGCAATCCAGAACTACAAATACATACACAGCAACACAGCAACTAAACAGTCA 780
 Db 870 TGTAGCAATCCAGAACTACAAATACATACACAGCAACACAGCAACTAAACAGTCA 929

QY 781 ACCAAGTCAAGTAATGACATGATGTCCTTGAACACAGCTCTACAACTGCTTTGAG 840
 Db 930 ACCAAGTCAAGTAATGACATGATGTCCTTGAACACAGCTCTACAACTGCTTTGAG 989
 QY 841 TCACGACATGTAGAACTGATGCGCTTGTCTTTGATCCAGCACAAAATCAAGTGAAC 900
 Db 990 TCACGACATGTAGAACTGATGCGCTTGTCTTTGATCCAGCACAAAATCAAGTGAAC 1049
 QY 901 ACCTGAGGTGTGCGATGCCACACGGAGATCATTCACCTTCATCCCTTACTCTCAAT 960
 Db 1050 ACCTGAGGTGTGCGATGCCACACGGAGATCATTCACCTTCATCCCTTACTCTCAAT 1109
 QY 961 GTCGTAAATGGAAGAACGATCGCTGTATTAATCCCTTCGTATCGTTCAACCATG 1020
 Db 1110 GTCGTAAATGGAAGAACGATCGCTGTATTAATCCCTTCGTATCGTTCAACCATG 1169
 QY 1021 GGTACCATGTCAGAGCCAGAACACCAAGTCCACACCGACTCCGGAACCTTAGTCCAG 1080
 Db 1170 GGTACCATGTCAGAGCCAGAACACCAAGTCCACACCGACTCCGGAACCTTAGTCCAG 1229
 QY 1081 CCGGCACTGCACCAATCTTAATAATAGATCAATCTTCTTGTGTTAGTCAGCTGCT 1140
 Db 1230 CCGGCACTGCACCAATCTTAATAATAGATCAATCTTCTTGTGTTAGTCAGCTGCT 1289
 QY 1141 ACGAAAGTTGGGGAAGATATGATTTCCAGAAAGAGGCACTCTCGTTAGTCTTTGC 1200
 Db 1290 ACGAAAGTTGGGGAAGATATGATTTCCAGAAAGAGGCACTCTCGTTAGTCTTTGC 1349
 QY 1201 GAAAGATTTACCATCTGAAACCTGTAAATCTTGAAAGCAAGTTATCAAAACAGAGAG 1260
 Db 1350 GAAAGATTTACCATCTGAAACCTGTAAATCTTGAAAGCAAGTTATCAAAACAGAGAG 1409
 QY 1261 TGTTCACACACTTTACTGCTTAAAGAAAGAAATGTTGCTCTCTCGTACCAAGAAATTTTA 1320
 Db 1410 TGTTCACACACTTTACTGCTTAAAGAAAGAAATGTTGCTCTCTCGTACCAAGAAATTTTA 1469
 QY 1321 TGATAAGCATATTAATCTGTAACTGAGGCTCAATAAGGCTTGTGNAATTAAGGGTGC 1380
 Db 1470 TGATAAGCATATTAATCTGTAACTGAGGCTCAATAAGGCTTGTGNAATTAAGGGTGC 1529
 QY 1381 TAATTCGATTTCCAGCCTTAGACCAATTAATTAAGACCTTGATGATGATCGACTAA 1440
 Db 1530 TAATTCGATTTCCAGCCTTAGACCAATTAATTAAGACCTTGATGATGATCGACTAA 1589
 QY 1441 TAAAGAAAAATGTGTGATGATTTATTTGGCATTCTTAGCAACCAATTAATCCATCCAGAGC 1500
 Db 1590 TAAAGAAAAATGTGTGATGATTTATTTGGCATTCTTAGCAACCAATTAATCCATCCAGAGC 1649
 QY 1501 ACTGGCAACCAAAATTTCTCAATTAAGTATGTAAGTGAAGCAAGTTCGATTTGCTCAAT 1560
 Db 1650 ACTGGCAACCAAAATTTCTCAATTAAGTATGTAAGTGAAGCAAGTTCGATTTGCTCAAT 1709
 QY 1561 AGCTGTAAGTATACACAGCTCAGATGTTATCAATTTTGTGATGAACATGATATATACAGTA 1620
 Db 1710 AGCTGTAAGTATACACAGCTCAGATGTTATCAATTTTGTGATGAACATGATATATACAGTA 1769
 QY 1621 TGAAGAGATGCAATATGTAACGCTCATATGCGCCATAGTCACTGATTTGGAAGATAG 1680
 Db 1770 TGAAGAGATGCAATATGTAACGCTCATATGCGCCATAGTCACTGATTTGGAAGATAG 1829
 QY 1681 CCGTTCTGTAAGAAAAAGTTGCAAGCTCAAGCCTTACTTAAGAAAAAGGATATCTTACC 1740
 Db 1830 CCGTTCTGTAAGAAAAAGTTGCAAGCTCAAGCCTTACTTAAGAAAAAGGATATCTTACC 1889
 QY 1741 TCCATCTCCAGACGAGATGTTAAAGCAATCCCACTGAGATAGTGCAGCACTATTTTA 1800
 Db 1890 TCCATCTCCAGACGAGATGTTAAAGCAATCCCACTGAGATAGTGCAGCACTATTTTA 1949
 QY 1801 CAATCTGTGAAAGGGGAAAAAGCAATTCACCTGTTTCACTTCCATATATGTTAGCA 1860
 Db 1950 CAATCTGTGAAAGGGGAAAAAGCAATTCACCTGTTTCACTTCCATATATGTTAGCA 2009

QY 1861 TACAGTTGAGGTTAAACCGGTAATTTGATTAATTCCTATAGAGATCAATTAATAT 1920
 Db 2010 TACAGTTGAGGTTAAACCGGTAATTTGATTAATTCCTATAGAGATCAATTAATAT 2069
 QY 1921 TAAATTTGCTGTTGATGATGATCACAATACAAAGCTCCAAATGSCATATACCTTGAAGA 1980
 Db 2070 TAAATTTGCTGTTGATGATGATCACAATACAAAGCTCCAAATGSCATATACCTTGAAGA 2129
 QY 1981 TTTGTTTGGCAGATTAAGTACTAGTGAACACCCCTGACCAAGCTCCACATTTCAATGA 2040
 Db 2130 TTTGTTTGGCAGATTAAGTACTAGTGAACACCCCTGACCAAGCTCCACATTTCAATGA 2189
 QY 2041 TGGATGGGCAATGCGCAGTGAAGCATGTGTTAGGCAAGAAAGACCAAGTGAAGATCCAA 2100
 Db 2190 TGGATGGGCAATGCGCAGTGAAGCATGTGTTAGGCAAGAAAGACCAAGTGAAGATCCAA 2249
 QY 2101 TAAAGATTTCAAAAGCGGATGAAGAGCCAGTGAAGAAACACCTGCTGAGCCAGAAATGCC 2160
 Db 2250 TAAAGATTTCAAAAGCGGATGAAGAGCCAGTGAAGAAACACCTGCTGAGCCAGAAATGCC 2309
 QY 2161 TCAAGTGAAGACTGAAAAAGTGAAGCCCACTCAAGAGAGAGAGATTTGCTTGGCA 2220
 Db 2310 TCAAGTGAAGACTGAAAAAGTGAAGCCCACTCAAGAGAGAGAGATTTGCTTGGCA 2369
 QY 2221 AGTAACGATTTCTAGTCTGAAGACCAATGCAACAGAACTAGCTGTTTACGAATTA 2280
 Db 2370 AGTAACGATTTCTAGTCTGAAGACCAATGCAACAGAACTAGCTGTTTACGAATTA 2429
 QY 2281 TTTGACTCTTCAATTAATGATTAACATATGATATGATGAGAGAGAGAGAGAGAGAGAGAG 2340
 Db 2430 TTTGACTCTTCAATTAATGATTAACATATGATATGATGAGAGAGAGAGAGAGAGAGAGAG 2489
 QY 2341 GTTGTAAAGAGATTAATCTTCTCATCTGTAAGTAAAGAAAAAATTAAC 2389
 Db 2490 GTTGTAAAGAGATTAATCTTCTCATCTGTAAGTAAAGAAAAAATTAAC 2538

RESULT 5
 AR218862
 LOCUS AR218862 8195 bp DNA linear PAT 25-SEP-2002
 DEFINITION Sequence 94 from patent US 6420135.
 ACCESSION AR218862
 VERSION AR218862.1 GI:23319796
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE
 1 (bases 1 to 8195)
 Kunisch, C.A., Choi, G.H., Dillon, P.S., Rosen, C.A., Barash, S.C.,
 Fannon, M.R., and Dougherty, B.A.,
 Streptococcus pneumoniae polynucleotides and sequences
 Patent: US 6420135-A 94 16-JUL-2002;
 TITLE JOURNAL
 FEATURES
 source location/Qualifiers
 1..8195
 /organism="unknown"
 BASE COUNT 2688 a 1622 c 1777 g 2105 t 3 others
 ORIGIN
 Query Match 100.0%; Score 2388; DB 6; Length 8195;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 TTCTTAAGAGTTGGACCTGTATCAAGCTTAGAACGTTAAGAAAAATATCGTTTCTTA 60
 Db 3053 TTCTTAAGAGTTGGACCTGTATCAAGCTTAGAACGTTAAGAAAAATATCGTTTCTTA 3112
 QY 61 TATAGATGAAAAAACAAGCAGCAAGCAAAAAAGGAGAAATTTGACTCTCGATGAGAGTAA 120
 Db 3113 TATAGATGAAAAAACAAGCAGCAAGCAAAAAAGGAGAAATTTGACTCTCGATGAGAGTAA 3172
 QY 121 GCGTGAAGAAATCAATGCTGAGCAAAATGCTCATCAAGATTAACAGCAAGGCTATGTAC 180
 Db 3173 GCGTGAAGAAATCAATGCTGAGCAAAATGCTCATCAAGATTAACAGCAAGGCTATGTAC 3232

181 TTCACATGGCGACCACTATCATTTATTAACAATGAGTTCCTTATGACGATCATCAG 240
3233 TTCAATGGCGACCACTATCATTTATTAACAATGAGTTCCTTATGACGATCATCAG 3292
241 TGAAGAATTACTCATGAAGAAGTCCAAACTATAGCTAAAAAGATGAGGATTTGTTAATGA 300
3293 TGAAGAATTACTCATGAAGAAGTCCAAACTATAGCTAAAAAGATGAGGATTTGTTAATGA 3352
301 GGTCAAGGTGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGATGC 360
3353 GGTCAAGGTGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGATGC 3412
361 TGCCACGCGGATTAAGTCCGATCAAAAGAGAAATCAATCGACAAAACAGAGCATAG 420
3413 TGCCACGCGGATTAAGTCCGATCAAAAGAGAAATCAATCGACAAAACAGAGCATAG 3472
421 TCAACATCGTGAAGGTGAATCTCAAGAAACGATGCTGTTGCCCTTGCGA 480
3473 TCAACATCGTGAAGGTGAATCTCAAGAAACGATGCTGTTGCCCTTGCGA 3532
481 AGGAGCGTATTAACAAGATGAGTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
3533 AGGAGCGTATTAACAAGATGAGTATATCTTTAATGCTTCTGATATCATAGAGATAC 3592
541 TGGTATGCTTATATCGTTCCTCATGAGATCATTAACATTAATCTTCAAGATGAGT 600
3593 TGGTATGCTTATATCGTTCCTCATGAGATCATTAACATTAATCTTCAAGATGAGT 3652
601 ATCAGCTAGCGAGTGGCTGCTGCGAAGACCTTCTATCTGCTGAGGAAATCTGTCAA 660
3653 ATCAGCTAGCGAGTGGCTGCTGCGAAGACCTTCTATCTGCTGAGGAAATCTGTCAA 3712
661 TTCAAGAACTTATCCCGCAAAATATAGCATTAACCTTAAGAAACAACTGGCTACCTTC 720
3713 TTCAAGAACTTATCCCGCAAAATATAGCATTAACCTTAAGAAACAACTGGCTACCTTC 3772
721 TGTAAAGCATCCAGGAATCAAAATATCTAACAAGCAACAGAGAACTTAACAGTCA 780
3773 TGTAAAGCATCCAGGAATCAAAATATCTAACAAGCAACAGAGAACTTAACAGTCA 3832
781 AGCAAGTCAAAAGTAAAGATGATGATGCTCTTGAAGACGCTCTAACAACCTGCTTGA 840
3833 AGCAAGTCAAAAGTAAAGATGATGATGCTCTTGAAGACGCTCTAACAACCTGCTTGA 3892
841 TCAAGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
3893 TCAAGCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3952
901 AGCTAGAGTGTGAGAGTGCACAGAGAGATCAATACCACTTACCTTACCTTCAAT 960
3953 AGCTAGAGTGTGAGAGTGCACAGAGAGATCAATACCACTTACCTTACCTTCAAT 4012
961 GTCTGAATTTGAAGAAAGCAATGCTGATTAATTCCTTCTGTTATGCTTCAACATG 1020
4013 GTCTGAATTTGAAGAAAGCAATGCTGATTAATTCCTTCTGTTATGCTTCAACATG 4072
1021 GGTACCAATTTCAAGGCCAGAAACCAAGTCCACACGACTCCGGAACCTAGTCAGG 1080
4073 GGTACCAATTTCAAGGCCAGAAACCAAGTCCACACGACTCCGGAACCTAGTCAGG 4132
1081 CCCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTTGGTATGACAGTGT 1140
4133 CCCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTTGGTATGACAGTGT 4192
1141 ACGAAAGTGGGAGAGATATGATTCGAGAGAAAGGACATCTCTGTTATGCTTTC 1200
4193 ACGAAAGTGGGAGAGATATGATTCGAGAGAAAGGACATCTCTGTTATGCTTTC 4252
1201 GAAAGATTTACATCTGAAAGCTGTTAAAAATCTTGAAGCAAGTTATCAAAAGAGAG 1260
4253 GAAAGATTTACATCTGAAAGCTGTTAAAAATCTTGAAGCAAGTTATCAAAAGAGAG 4312

1261 TGTTCACACACTTTAATCTGTAAGAAAAAGAAATGTTGCTCTCTGTCGACAAAGATTTTA 1320
4313 TGTTCACACACTTTAATCTGTAAGAAAAAGAAATGTTGCTCTCTGTCGACAAAGATTTTA 4372
1321 TGTAAAGCATATATCTGTTAATCTGAGGCTCATTAAGCCTTGTGTTAAATTAAGGCTG 1380
4373 TGTAAAGCATATATCTGTTAATCTGAGGCTCATTAAGCCTTGTGTTAAATTAAGGCTG 4432
1381 TAAATTCGATTTCCAGCCTTGAACAAATTTTAAAGAGCTTGAATGATGATGACTTA 1440
4433 TAAATTCGATTTCCAGCCTTGAACAAATTTTAAAGAGCTTGAATGATGATGACTTA 4492
1441 TAAAGAAAAATTTGATAGATATTAATTTGACCTTCAACCAATTAACCAATCAAGAG 1500
4493 TAAAGAAAAATTTGATAGATATTAATTTGACCTTCAACCAATTAACCAATCAAGAG 4552
1501 ACTTGGCAACCAAAATTTCTCAAAATTTGAGTATCTGAAGACAAAGTTGCTATGCTCAAT 1560
4553 ACTTGGCAACCAAAATTTCTCAAAATTTGAGTATCTGAAGACAAAGTTGCTATGCTCAAT 4612
1561 AGCTGATAGATTAACAGCTGAGATGTTTATGATGATGATGATGATGATGATGATGATG 1620
4613 AGCTGATAGATTAACAGCTGAGATGTTTATGATGATGATGATGATGATGATGATGATG 4672
1621 TGAAGAGATGATATGATTAACAGCTGATGAGGCTATGATGATGATGATGATGATGATG 1680
4673 TGAAGAGATGATATGATTAACAGCTGATGAGGCTATGATGATGATGATGATGATGATG 4732
1681 CTTTCTGATTAAGAAAAAGTTGCACTCAAGCTTATCTTAAGAAAAAGTATCTTACC 1740
4733 CTTTCTGATTAAGAAAAAGTTGCACTCAAGCTTATCTTAAGAAAAAGTATCTTACC 4792
1741 TCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGAGATGATGAGAGAGATTTTA 1800
4793 TCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGAGATGATGAGAGAGATTTTA 4852
1801 CAATCTGTGTAAGAGGAAAAAGCAATTCATCTGCTTCACTTCAATATATGTTGAGCA 1860
4853 CAATCTGTGTAAGAGGAAAAAGCAATTCATCTGCTTCACTTCAATATATGTTGAGCA 4912
1861 TACAGTGAAGTTAAAAAGGTAATTTGATTAATTCCTCATTAAGATCAATTAATAT 1920
4913 TACAGTGAAGTTAAAAAGGTAATTTGATTAATTCCTCATTAAGATCAATTAATAT 4972
1921 TAAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1980
4973 TAAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 5032
1981 TTTGTTTGCAGATTAAGTACTAGTAAACACCTGACGAACTGCACATTCCTAAATGA 2040
5033 TTTGTTTGCAGATTAAGTACTAGTAAACACCTGACGAACTGCACATTCCTAAATGA 5092
2041 TGAATGGGCAATTCAGTGAAGATGTTGTTAGGCAAGAAACCAACAGTGAAGTCCAA 2100
5093 TGAATGGGCAATTCAGTGAAGATGTTGTTAGGCAAGAAACCAACAGTGAAGTCCAA 5152
2101 TAAAGACTTCAAGCGAGTGAAGAGCCAGTGAAGAAACCTGCTGAGCCAGAGTCC 2160
5153 TAAAGACTTCAAGCGAGTGAAGAGCCAGTGAAGAAACCTGCTGAGCCAGAGTCC 5212
2161 TCAAGTGAAGACTGAAGAAAGTGAAGCCCAACTCAAAAGACAGAAAGTTTCTTGGCAA 2220
5213 TCAAGTGAAGACTGAAGAAAGTGAAGCCCAACTCAAAAGACAGAAAGTTTCTTGGCAA 5272
2221 AGTAAAGGATTTAGTCTGAAGCCATGCAACGAAATCTTACCTGTTTACGAATTA 2280
5273 AGTAAAGGATTTAGTCTGAAGCCATGCAACGAAATCTTACCTGTTTACGAATTA 5332
2281 TTTGACTCTTCAATTTATGATTAACAATAGTATATGATGATGATGATGATGATGATG 2340
5333 TTTGACTCTTCAATTTATGATTAACAATAGTATATGATGATGATGATGATGATGATG 5392
2341 GTTGTAAAGAGATATCTCTCATCTGTAAGTAAAGAAAAATTAAC 2389

Db 5393 GTTGTAAAGAAAGTAATCTTCATCTGTAGTAAGAAATAAATTAAC 5441

RESULT 6
BD003774 8195 bp DNA linear PAT 31-JAN-2002

LOCUS BD003774 Polynucleotide of Streptococcus pneumoniae and sequence.
ACCESSION BD003774.1 GI:18631735
VERSION JP 2001501833-A/94.
KEYWORDS unclassified
SOURCE unclassified
ORGANISM unclassified
REFERENCE 1 (bases 1 to 8195)
AUTHORS Kunsch, C.A., Choi, G.H., Dillon, P.J., Rosen, C.A., Bara, S.C.,
Fannon, M. and Dougherty, B.A.
TITLE Polynucleotide of Streptococcus pneumoniae and sequence
JOURNAL Patent: JP 2001501833-A 94 13-FEB-2001;
HUMAN GENOME SCIENCES INC

COMMENT
OS Unidentified
PN JP 2001501833-A/94
PD 13-FEB-2001
PF 30-OCT-1997 JP 1998520718
PR 31-OCT-1996 US 60/029960
PI CHARLES A KUNSCH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, PI
STEVEN C BARASH,
PI MICHAEL FANNON, BRIAN A DOUGHERTY
PC C12N15/09, A01K67/027, C07K14/315, C07K16/12, C12N1/15, C12N1/19,
PC C12N1/21,
PC C12N5/10, C12P21/02, C12Q1/68, G06F17/30, C12N15/00, C12N5/00, PC
G06F15/40
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1. 8195
FT Location/Qualifiers
1. 8195
/organism="unclassified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

BASE COUNT 2688 a 1622 c 1777 g 2105 t 3 others

ORIGIN

Query Match 100.0%; Score 2388; DB 6; Length 8195;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGAGCTGATCAAGCTAGAACGTTTAAAGAAATTAATGTTTCTTA 60
Db 3053 TTCTTACGAGTTGGAGCTGATCAAGCTAGAACGTTTAAAGAAATTAATGTTTCTTA 3112
QY 61 TATAGATGAAAAAACAAGCGACGCAAAAAAAGGAAATTTGACTCTCGATGAGGTTAGCA 120
Db 3113 TATAGATGAAAAAACAAGCGACGCAAAAAAAGGAAATTTGACTCTCGATGAGGTTAGCA 3172
QY 121 GCGTGAAGAAATTAATGCTGAGCAAAATCGTCATCAAGATACAGACCAAGGCTATGTCAC 180
Db 3173 GCGTGAAGAAATTAATGCTGAGCAAAATCGTCATCAAGATACAGACCAAGGCTATGTCAC 3232
QY 181 TTCACATGGGACCACTATCATTAATTAACAATGTTAAGGTTCCCTTATGAGGCTATCATGAC 240
Db 3233 TTCACATGGGACCACTATCATTAATTAACAATGTTAAGGTTCCCTTATGAGGCTATCATGAC 3292
QY 241 TGAAGAATTAATCATGATAAAGATCAAACTATAAAGCTTAAAGATGAGGATATTGTTAATGA 300
Db 3293 TGAAGAATTAATCATGATAAAGATCAAACTATAAAGCTTAAAGATGAGGATATTGTTAATGA 3352
QY 301 GGTCAAGGTTGATGTTATCAAGGTAATGAAATTAATGTTTAAAGATGAC 360
Db 3353 GGTCAAGGTTGATGTTATCAAGGTAATGAAATTAATGTTTAAAGATGAC 3412

QY 361 TGCCACGCGGATTAACGTCCTGATCAAAAGAGAAATCAATGCACAAAAACAAGAGCATAG 420
Db 3413 TGCCACGCGGATTAACGTCCTGATCAAAAGAGAAATCAATGCACAAAAACAAGAGCATAG 3472
QY 421 TCACATCGTGAAGTGGAACTTCAGAAAGCATGCTGTTGCTTGGACGTTTGGCA 480
Db 3473 TCACATCGTGAAGTGGAACTTCAGAAAGCATGCTGTTGCTTGGACGTTTGGCA 3532
QY 481 AGGAGCTTACTACAGATGATGTTATCTTTAATGCTTGTGATCATGAGGATAC 540
Db 3533 AGGAGCTTACTACAGATGATGTTATCTTTAATGCTTGTGATCATGAGGATAC 3592
QY 541 TGGATGCTTATATGCTTCTCATGAGATCATTAACATTAATCTTAAGATGAGTT 600
Db 3593 TGGATGCTTATATGCTTCTCATGAGATCATTAACATTAATCTTAAGATGAGTT 3652
QY 601 ATCAGTACGAGTTGGCTGCTGACAGAGCCTTCTATCTGTCGAGAAATCTGTCAAA 660
Db 3653 ATCAGTACGAGTTGGCTGCTGACAGAGCCTTCTATCTGTCGAGAAATCTGTCAAA 3712
QY 661 TTCAGAACCTATGCGCGCAAAATAGAGATTAACCTCAAGAACAACTGGGTACTTC 720
Db 3713 TTCAGAACCTATGCGCGCAAAATAGAGATTAACCTCAAGAACAACTGGGTACTTC 3772
QY 721 TGTAAAGCAATCCAGAACTACAAATTAACAACAAGCAACAACAGCACTAACAGTCA 780
Db 3773 TGTAAAGCAATCCAGAACTACAAATTAACAACAAGCAACAACAGCACTAACAGTCA 3832
QY 781 AGCAAGTCAAGTAATGACATGATGCTCTTGAACAGCTCTCAAACTGCTTTGAG 840
Db 3833 AGCAAGTCAAGTAATGACATGATGCTCTTGAACAGCTCTCAAACTGCTTTGAG 3892
QY 841 TCAAGCACTATGAGATCTGATGCGCTTCTTGTATCCAGACAAATACAAAGTGAAC 900
Db 3893 TCAAGCACTATGAGATCTGATGCGCTTCTTGTATCCAGACAAATACAAAGTGAAC 3952
QY 901 ACTAGAGGTTGAGAGTCCACAGAGATCATTAACCTTATCCCTTACTCTCAAT 960
Db 3953 ACTAGAGGTTGAGAGTCCACAGAGATCATTAACCTTATCCCTTACTCTCAAT 4012
QY 961 GTCTGATTTGAAGAACGATGCTGCTGTTATTCCTCTTGTGTTCAACCATTTG 1020
Db 4013 GTCTGATTTGAAGAACGATGCTGCTGTTATTCCTCTTGTGTTCAACCATTTG 4072
QY 1021 GGTACAGATTAAGGCCCAACCAACCAAGTCCACCAACCGACTCCGGAACCTAGTCAAG 1080
Db 4073 GGTACAGATTAAGGCCCAACCAACCAAGTCCACCAACCGACTCCGGAACCTAGTCAAG 4132
QY 1081 CCCGCACTGACCAAAATCTTAATAATGACTCAAAATCTTCTTGGTTAGTCAAGCTGGT 1140
Db 4133 CCCGCACTGACCAAAATCTTAATAATGACTCAAAATCTTCTTGGTTAGTCAAGCTGGT 4192
QY 1141 ACGAAAGTTGGGGAAGATATGATTCGAAAGAAAAGGCAATCTCGTTATGTCCTTTGC 1200
Db 4193 ACGAAAGTTGGGGAAGATATGATTCGAAAGAAAAGGCAATCTCGTTATGTCCTTTGC 4252
QY 1201 GAAAGTTTACATCTGAAACGTTTAAATCTTGAAGCAAGTATCAAAACAAGAGG 1260
Db 4253 GAAAGTTTACATCTGAAACGTTTAAATCTTGAAGCAAGTATCAAAACAAGAGG 4312
QY 1261 TGTTCACACACTTAACTGCTTAAAGAAAGAAATGTCCTCGTGAACAAGAAATTTTA 1320
Db 4313 TGTTCACACACTTAACTGCTTAAAGAAAGAAATGTCCTCGTGAACAAGAAATTTTA 4372
QY 1321 TGAATAAGCATTAATCTGTTAACTGAGGCTCATAAAGCCTTGTGNAATTAAGGCTG 1380
Db 4373 TGAATAAGCATTAATCTGTTAACTGAGGCTCATAAAGCCTTGTGNAATTAAGGCTG 4432
QY 1381 TAATTCGATTTCCAAAGCCTTGAACAATTAATGAAGCCTTGAATGAATTCACATA 1440
Db 4433 TAATTCGATTTCCAAAGCCTTGAACAATTAATGAAGCCTTGAATGAATTCACATA 4492
QY 1441 TAAAGAAAATGTAGATGATTTATTTGACATTCCTAGACCAATTAACCATTCAGAGG 1500

```

Db      4493  TAAAGAAAAATGTAGATGATTTATGTGCAATTCCTAGCAGCAATTAACCATCCAGAGG 4552
Oy      1501  ACTTGCAACCAATTCCTCAATGAGATATCTGAAGCGAAGTCCGATGCTCANT 1560
Db      4553  ACTTGCAACCAATTCCTCAATGAGATATCTGAAGCGAAGTCCGATGCTCANT 4612
Oy      1561  AGCTGATAGATATACACCTGAGATGATGATTTTGTGTAACATGATATATAGTGA 1620
Db      4613  AGCTGATAGATATACACCTGAGATGATGATTTTGTGTAACATGATATATAGTGA 4672
Oy      1621  TGAAGAGATGATATGTAAGCCTCAATGAGGCAATGATGATGATGAAAAAGATG 1680
Db      4673  TGAAGAGATGATATGTAAGCCTCAATGAGGCAATGATGATGATGAAAAAGATG 4732
Oy      1681  CTTTCTGTAAGAAAAAGTTCGAGCTCAACCTATATATAAGAAAAAGATGCTTACC 1740
Db      4733  CTTTCTGTAAGAAAAAGTTCGAGCTCAACCTATATATAAGAAAAAGATGCTTACC 4792
Oy      1741  TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGAGATGAGCAGCTATTTA 1800
Db      4793  TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGAGATGAGCAGCTATTTA 4852
Oy      1801  CAATGCTGTAAGAAAAAGGAAAAAGCAATTCCTGCTGCTGCTGCTGCTGCTGCTG 1860
Db      4853  CAATGCTGTAAGAAAAAGGAAAAAGCAATTCCTGCTGCTGCTGCTGCTGCTGCTG 4912
Oy      1861  TACAGTTGAGTTAAAGAGGTAATTTGATTTTCTCTCTAAGAGATGATTAATATAT 1920
Db      4913  TACAGTTGAGTTAAAGAGGTAATTTGATTTTCTCTCTAAGAGATGATTAATATAT 4972
Oy      1921  TAAATTTGCTGTTGATGATCACAATCAATCAATCAATCAATCAATCAATCAATCA 1980
Db      4973  TAAATTTGCTGTTGATGATCACAATCAATCAATCAATCAATCAATCAATCAATCA 5032
Oy      1981  TTTGTTTGCGAGATTAAGTACTAGTAGAACAACCTGAGCAAGCTGCAATTTCAATGA 2040
Db      5033  TTTGTTTGCGAGATTAAGTACTAGTAGAACAACCTGAGCAAGCTGCAATTTCAATGA 5092
Oy      2041  TGGATGGGGCAATGCGAGTGAAGATGTTTGAAGCAAGAAAGCAAGTGAAGATCCAA 2100
Db      5093  TGGATGGGGCAATGCGAGTGAAGATGTTTGAAGCAAGAAAGCAAGTGAAGATCCAA 5152
Oy      2101  TGAAGACTTCAAGCGAGTGAAGAGCAGTAGAGAGAAACACTGCTGAGCCGAGAGTCCC 2160
Db      5153  TGAAGACTTCAAGCGAGTGAAGAGCAGTAGAGAGAAACACTGCTGAGCCGAGAGTCCC 5212
Oy      2161  TCAAGTAGAGACTGAAAAAGTGAAGAGCCCACTCAAGAGAGAGAGTCTTGGCA 2220
Db      5213  TCAAGTAGAGACTGAAAAAGTGAAGAGCCCACTCAAGAGAGAGAGTCTTGGCA 5272
Oy      2221  AGTACGAGATTTAGTCTGAAAGCCCAATGCAACAGAACTCTAGCTGTTTACGAATTA 2280
Db      5273  AGTACGAGATTTAGTCTGAAAGCCCAATGCAACAGAACTCTAGCTGTTTACGAATTA 5332
Oy      2281  TTTGACTCTTCAATTTATGATGATACATATGATGATGATGATGATGATGATGATG 2340
Db      5333  TTTGACTCTTCAATTTATGATGATACATATGATGATGATGATGATGATGATGATG 5392
Oy      2341  GTTGTAAAGAGAGTAACTCTTCACTGTTAAGTAAAGAAAAATTAAC 2389
Db      5393  GTTGTAAAGAGAGTAACTCTTCACTGTTAAGTAAAGAAAAATTAAC 5441

```

```

RESULT 7
AE007418/c 10256 bp DNA linear BCT 31-AUG-2001
LOCUS      Streptococcus pneumoniae TIGR4 section 101 of 194 of the complete
DEFINITION genome.
ACCESSION  AE007418 AE005672
VERSION     AE007418.1 GI:14972649
KEYWORDS    Streptococcus pneumoniae TIGR4
SOURCE

```

```

ORGANISM Streptococcus pneumoniae TIGR4
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 10256)
AUTHORS Tetteijn, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D.,
Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J.,
Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D.,
Umayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D.,
Holtzapple, E., Khouri, H., Wolf, A.M., Uitterlinden, T.R., McDonald, L.A.,
McDonald, L.A., Feldblyum, T.V., Angiuoli, S., Dickinson, T.,
Hickey, E.K., Holt, I.E., Loftus, B.J., Yang, F., Smith, H.O.,
Venter, J.C., Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and
Fraser, C.M.
Complete genome sequence of a virulent isolate of Streptococcus
pneumoniae
JOURNAL Science 293 (5529), 498-506 (2001)
MEDLINE 21357209
PUBMED 11463916
REFERENCE 2 (bases 1 to 10256)
AUTHORS Tetteijn, H., Nelson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D.,
Peterson, S., Heidelberg, J., DeBoy, R.T., Haft, D.H., Dodson, R.J.,
Durkin, A.S., Gwinn, M., Kolonay, J.F., Nelson, W.C., Peterson, J.D.,
Umayam, L.A., White, O., Lewis, M.R., Radune, D., Holtzapple, E.,
Khouri, H., Wolf, A.M., Uitterlinden, T.R., Hansen, C.L., McDonald, L.A.,
Feldblyum, T.V., Angiuoli, S., Gesuwan, P., Hickey, E.K., Holt, I.E.,
Loftus, B.J., Ujwal, M.L., Yang, F., Smith, H.O., Venter, J.C.,
Dougherty, B.A., Morrison, D.A., Hollingshead, S.K. and Fraser, C.M.
Direct Submission
Submitted (29-JUN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
location/Qualifiers
1..10256
/organism="Streptococcus pneumoniae TIGR4"
/mol_type="genomic DNA"
/strain="TIGR4"
/db_xref="taxon:170187"
/complement(100..702)
/gene="Spi170"
/complement(100..702)
/gene="Spi170"
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAK75279.1"
/db_xref="GI:14972650"
/translation="MSEVDNFEAVNEFTSDTQANSIYQSLFKFPDKKPFSGDLIF
TWKSPSLVKEGDYIGRDSQVDNLRVGNLFPNLTNRKXSLNNNRGCGWDFPHDF
DIYLDHAKYAYEQKNVNIKEYPLKRAILHOENALYRFRSPNDFLEQNYLKTIIQ
VSKETPSEMDPNFNKISERILFERGSKMLNDLKSNNYK"
/complement(760..1473)
/gene="Spi171"
/complement(760..1473)
/feature="identified by match to PFM protein family HMM
PF00702"
/codon_start=1
/transl_table=1
/product="hydrolase, haloacid dehalogenase-like family"
/protein_id="AAK75280.1"
/db_xref="GI:14972651"
/translation="MEYKFLFLFDLHTLDPDAEDVALTQLKEBQVADIQAYKDY
VPMKALWKDLELEKISKEQVNNRFRSLFAHQEQEDEFQAGRYOYFLAOGQSTG
GAHDLDSLIERDYNLYAANGITAIOTGSLAOSGLAPVNOYPIESLOTORDALE
YKIQIAGFSKERTIMIGDSLADIGGNAGIDITWNPPLHNHTQAQPIYEV
SYDLDLCLKNLEKITP"
/complement(1174..1843)
/gene="Spi172"
/complement(1174..1843)
/gene="Spi172"
/feature="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1

```


Oy 601 ATCACTAGCGAGTGGCTGCTGCGAAGACCTTCTATCTGGTCGAGGAATCTGTCAA 660
 Db 7137 ATCACTAGCGAGTGGCTGCTGCGAAGACCTTCTATCTGGTCGAGGAATCTGTCAA 7078
 Oy 661 TTCAAGAACTATGCGCCGACAAATAGCGATTAACCTTCAAGAACAACTGGGATCCTTC 720
 Db 7077 TTCAAGAACTATGCGCCGACAAATAGCGATTAACCTTCAAGAACAACTGGGATCCTTC 7018
 Oy 721 TGTAGCAATCCAGAACTACAAATCTAACAAGCAAGCAACACCAACCTAAGCA 780
 Db 7017 TGTAGCAATCCAGAACTACAAATCTAACAAGCAAGCAACACCAACCTAAGCA 6958
 Oy 781 AGCAAGTAAAGTAATGACATGATAGTCTCTGAAACAGCTCTCAAACTGCCCTTGGAG 840
 Db 6957 AGCAAGTAAAGTAATGACATGATAGTCTCTGAAACAGCTCTCAAACTGCCCTTGGAG 6898
 Oy 841 TCAACGACATGTAGATCTGATGGCTTGTCTTTGATCCAGACAAATCACAAGTCGAAC 900
 Db 6897 TCAACGACATGTAGATCTGATGGCTTGTCTTTGATCCAGACAAATCACAAGTCGAAC 6838
 Oy 901 AGCTAGAGTGTGTCAGTGCACACGAGATCATTAACCACTTACCTTACTCTCAAT 960
 Db 6837 AGCTAGAGTGTGTCAGTGCACACGAGATCATTAACCACTTACCTTACTCTCAAT 6778
 Oy 961 GTCTGAATTTGAAAGCAATGCTCGTATTAATTCCTCCCTGCTTATCGTTCAAACTTG 1020
 Db 6777 GTCTGAATTTGAAAGCAATGCTCGTATTAATTCCTCCCTGCTTATCGTTCAAACTTG 6718
 Oy 1021 GGTACCAATTCAGAGCCAGAACCAAGTCACACCCAGCTCCGGAACCTAGTCAGG 1080
 Db 6717 GGTACCAATTCAGAGCCAGAACCAAGTCACACCCAGCTCCGGAACCTAGTCAGG 6658
 Oy 1081 CCCGACCTGACCAAACTTTAAATAGACTCAAACTTCTTTGTTAGTCACTGCT 1140
 Db 6657 CCCGACCTGACCAAACTTTAAATAGACTCAAACTTCTTTGTTAGTCACTGCT 6598
 Oy 1141 ACGAAAGTTGGGGAAGATATGTTTCGAAGAAAGGCACTCTCGTTATGTCTTTC 1200
 Db 6597 ACGAAAGTTGGGGAAGATATGTTTCGAAGAAAGGCACTCTCGTTATGTCTTTC 6538
 Oy 1201 GAAAGATTACCATCTGAACCTGTTAAATCTTGAAGCAAGTTATCAAAACAAGAG 1260
 Db 6537 GAAAGATTACCATCTGAACCTGTTAAATCTTGAAGCAAGTTATCAAAACAAGAG 6478
 Oy 1261 TGTTCACACACTTAACTGCTAAAAAAGAAATGTGCTCTCGTGCACAAAGATTTTA 1320
 Db 6477 TGTTCACACACTTAACTGCTAAAAAAGAAATGTGCTCTCGTGCACAAAGATTTTA 6418
 Oy 1321 TGATTAAGCATTAATCTGTTAACTGAGGCTCATTAAGGCTGTTGNAATTAAGGTCG 1380
 Db 6417 TGATTAAGCATTAATCTGTTAACTGAGGCTCATTAAGGCTGTTGNAATTAAGGTCG 6358
 Oy 1381 TAATTCGATTTCCAGGCTTAGACAAATTAATTAAGCGCTGAATGAATGACATTA 1440
 Db 6357 TAATTCGATTTCCAGGCTTAGACAAATTAATTAAGCGCTGAATGAATGACATTA 6298
 Oy 1441 TAAAGAAAAATTTGGTAGATTTATTTGGCATTTCTAGACCAATTAACCCATCCAGAGC 1500
 Db 6297 TAAAGAAAAATTTGGTAGATTTATTTGGCATTTCTAGACCAATTAACCCATCCAGAGC 6238
 Oy 1501 ACTGGCAAAACCAATTTCTCAATAGATATCTGAAGCGAAGTTCGATTTGCTCAAT 1560
 Db 6237 ACTGGCAAAACCAATTTCTCAATAGATATCTGAAGCGAAGTTCGATTTGCTCAAT 6178
 Oy 1561 AGCTGATAGTATCAACGTCAGATGTTTATCTTTGATGAACATGATATATCAATGA 1620
 Db 6177 AGCTGATAGTATCAACGTCAGATGTTTATCTTTGATGAACATGATATATCAATGA 6118
 Oy 1621 TGAAGAGATGATATTTGAAGCCCTCATTTGGGCAATGCTGCTGATTTGAAAAAGATG 1680
 Db 6117 TGAAGAGATGATATTTGAAGCCCTCATTTGGGCAATGCTGCTGATTTGAAAAAGATG 6058

Oy 1681 CTTTCTGATAGGAAAAAGTTGACGCTCAAGCCTATACTAAGAAAAAGATATCTACC 1740
 Db 6057 CTTTCTGATAGGAAAAAGTTGACGCTCAAGCCTATACTAAGAAAAAGATATCTACC 5998
 Oy 1741 TCCATCTCCAGACGCAATGTTAAAGCAATTCCAACTGAGATATGTCAGACGATTTTA 1800
 Db 5997 TCCATCTCCAGACGCAATGTTAAAGCAATTCCAACTGAGATATGTCAGACGATTTTA 5938
 Oy 1801 CAATCGTGTGAAGGGGAAAAAGAAATTCACCTGCTCGAATTCATATATGTTGAGCA 1860
 Db 5937 CAATCGTGTGAAGGGGAAAAAGAAATTCACCTGCTCGAATTCATATATGTTGAGCA 5878
 Oy 1861 TACAGTTAGGTTAAAAACGTAATTTGATTAATTCCTATPAAGATCATTAATAT 1920
 Db 5877 TACAGTTAGGTTAAAAACGTAATTTGATTAATTCCTATPAAGATCATTAATAT 5818
 Oy 1921 TAAATTTGCTTTGTTTATGATATCAACATACAAAGCTCCAAATGCTATACCTTGAAGA 1980
 Db 5817 TAAATTTGCTTTGTTTATGATATCAACATACAAAGCTCCAAATGCTATACCTTGAAGA 5758
 Oy 1981 TTTGTTGGCGACATTAAGTACTACGTAGAACACCCGAGCAAGCTCACATTTTAATGA 2040
 Db 5757 TTTGTTGGCGACATTAAGTACTACGTAGAACACCCGAGCAAGCTCACATTTTAATGA 5698
 Oy 2041 TGGATGGGGCAATGTCAGTGCATGTGTTAGGCAAGAAACCAACAGTAAGATCCAAA 2100
 Db 5697 TGGATGGGGCAATGTCAGTGCATGTGTTAGGCAAGAAACCAACAGTAAGATCCAAA 5638
 Oy 2101 TAAACCTTCAAGCCGATGAGAGCCAGTGAAGAAACCTGCTAGCCAGAAATGCC 2160
 Db 5637 TAAACCTTCAAGCCGATGAGAGCCAGTGAAGAAACCTGCTAGCCAGAAATGCC 5578
 Oy 2161 TCAAGTAGAGCTGAAAAAGTAGAAGCCCAACTCAAAAGAGCAAGATTTGCTTGGCAA 2220
 Db 5577 TCAAGTAGAGCTGAAAAAGTAGAAGCCCAACTCAAAAGAGCAAGATTTGCTTGGCAA 5518
 Oy 2221 AGTAACGGAATCTAGTCTGAAGCCCAATGCAACAGAACTCTAGCTGTTACGAATAA 2280
 Db 5517 AGTAACGGAATCTAGTCTGAAGCCCAATGCAACAGAACTCTAGCTGTTACGAATAA 5458
 Oy 2281 TTTGACTCTTCAATTTATGATTAACATATGATATATGTCAGACAGAAAAATTAATTCTTC 2340
 Db 5457 TTTGACTCTTCAATTTATGATTAACATATGATATATGTCAGACAGAAAAATTAATTCTTC 5398
 Oy 2341 GTTGTAAAAAGAGTAATCCTTCATCTGTAAGTAAGAAAAATTAAC 2389
 Db 5397 GTTGTAAAAAGAGTAATCCTTCATCTGTAAGTAAGAAAAATTAAC 5349

RESULT 8
 AX571763/c
 LOCUS AX571763 349980 bp DNA Linear PAT 29-NOV-2002
 DEFINITION Sequence 4982 from Patent WO02077021.
 ACCESSION AX571763
 VERSION AX571763.1 GI:26003955
 KEYWORDS
 SOURCE
 ORGANISM
 Streptococcus pneumoniae
 Streptococcus pneumoniae
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 REFERENCE
 1 Maeligani, V., Tettelin, H. and Fraser, C.
 AUTHORS Streptococcus pneumoniae proteins and nucleic acids
 TITLE Patent: WO 02077021-A 4982 03-OCT-2002.
 JOURNAL Chiron Spa (IT) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
 FEATURES
 source
 1. 349980
 /organism="Streptococcus pneumoniae"
 /mol_type="genomic DNA"
 /db_xref="taxon:1313"
 /note="seq 4979 too long: 2,162,598 bases-replaced by
 following seq:-seq 4979: from 0.000,001 to 0.349,980-seq
 4980: from 0.300,001 to 0.649,980-seq 4981: from 0.600,001

to 0.949,980-seq 4982: from 0.900,001 to 1.249,980-seq
 4983: from 1.200,001 to 1.549,980-seq 4984: from 1.500,001
 to 1.849,980-seq 4985: from 1.800,001 to 2.149,980-seq
 4986: from 2.100,001 to 2.162,598"
 BASE COUNT 106896 a 71720 c 64704 g 106660 t
 ORIGIN

Query Match 100.0%; Score 2388; DB 6; Length 349980;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAACGTTAAGGAAATPACGTGTTTCTTA 60
 DB 106950 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAACGTTAAGGAAATPACGTGTTTCTTA 106891

QY 61 TATAGATGAAAAACAAGCGACGAAAAACGGGAATTTTACCTCCGTAGAGTTAGCA 120
 DB 106890 TATAGATGAAAAACAAGCGACGAAAAACGGGAATTTTACCTCCGTAGAGTTAGCA 106831

QY 121 GCGTGAAGAAATCAATGCTGAGCAAAATCGTCATCAAGATPACAGACCAAGGCTATGTCAC 180
 DB 106830 GCGTGAAGAAATCAATGCTGAGCAAAATCGTCATCAAGATPACAGACCAAGGCTATGTCAC 106771

QY 181 TTCACATGAGCGACCACTATCATTTATTACAATGGTAAAGTTCTTATGACGCTATCATCAG 240
 DB 106770 TTCACATGAGCGACCACTATCATTTATTACAATGGTAAAGTTCTTATGACGCTATCATCAG 106711

QY 241 TGAAGAATTAATCATGATAAAGATCCAAACTTAAGCTTAAAGATGAGAGATTTGTTAATGA 300
 DB 106710 TGAAGAATTAATCATGATAAAGATCCAAACTTAAGCTTAAAGATGAGAGATTTGTTAATGA 106651

QY 301 GGTCAAGGCTGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAGGATGC 360
 DB 106650 GGTCAAGGCTGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAGGATGC 106591

QY 361 TGGCCACGGCGATACGTCCTGTCACAAAGAGAAATCAATCGACAAACCAAGAGCATAG 420
 DB 106590 TGGCCACGGCGATACGTCCTGTCACAAAGAGAAATCAATCGACAAACCAAGAGCATAG 106531

QY 421 TCACACTCGTGAAGGTGGAACCTCCAGAAACCATGCTGCTTGGCTGGACGTTGCGA 480
 DB 106530 TCACACTCGTGAAGGTGGAACCTCCAGAAACCATGCTGCTTGGCTGGACGTTGCGA 106471

QY 481 AGGACGCTATCTACAGATGATGGTATATCTTTAATGCTTCGATATCATAGAGATAC 540
 DB 106470 AGGACGCTATCTACAGATGATGGTATATCTTTAATGCTTCGATATCATAGAGATAC 106411

QY 541 TGGTATGCTTATATCGTTCCTCATGAGATCATTAACATTCATTCTTAAGATGATT 600
 DB 106410 TGGTATGCTTATATCGTTCCTCATGAGATCATTAACATTCATTCTTAAGATGATT 106351

QY 601 ATCAGCTAGCGAGTTGGCTGCTGCAAGAGCTTCATGCTGGTGGAGAAATTTGTCAA 660
 DB 106350 ATCAGCTAGCGAGTTGGCTGCTGCAAGAGCTTCATGCTGGTGGAGAAATTTGTCAA 106291

QY 661 TTCAGAACCTATCGCGCAAAATAAGCGATTAACATTCAGAGAAACCACTGGGTACTCTTC 720
 DB 106290 TTCAGAACCTATCGCGCAAAATAAGCGATTAACATTCAGAGAAACCACTGGGTACTCTTC 106231

QY 721 TGTAAACAATCCAGAACTACAAATACTAACACAGCAACAGAGCACTAACAGTCA 780
 DB 106230 TGTAAACAATCCAGAACTACAAATACTAACACAGCAACAGAGCACTAACAGTCA 106171

QY 781 AGCAAGTCAAAGTATGACATTGATAGTCTCTTGAACAGCTCTTACAACTGCTTTGAG 840
 DB 106170 AGCAAGTCAAAGTATGACATTGATAGTCTCTTGAACAGCTCTTACAACTGCTTTGAG 106111

QY 841 TCAAGACATGTGAATCTGATGGCTGTCTTGTATGACGACCAATACAAAGTGAAC 900
 DB 106110 TCAAGACATGTGAATCTGATGGCTGTCTTGTATGACGACCAATACAAAGTGAAC 106051

QY 901 AGTAGAGGTGTGCAAGTGCACACGAGATCATTTACATTGATCCCTTACTCTCAAT 960
 DB 106051 AGTAGAGGTGTGCAAGTGCACACGAGATCATTTACATTGATCCCTTACTCTCAAT 104991

DB 106050 AGTAGAGGTGTGCAAGTGCACACGAGATCATTTACCACTTATCCCTTACTCTCAAT 105991

QY 961 GTCCTAATTTGAGAGAACGAATGCTGCTATTTACCTGCTTATCGTTAAACATG 1020
 DB 105990 GTCCTAATTTGAGAGAACGAATGCTGCTATTTACCTGCTTATCGTTAAACATG 105931

QY 1021 GGTACAGATTCAGAGCGAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCCAG 1080
 DB 105930 GGTACAGATTCAGAGCGAGAACCAACCAAGTCCACACCGACTCCGGAACCTAGTCCAG 105871

QY 1081 CCCGCACTGACCAAAATCTTAAATAGACTCAAAATCTCTTGTGTTAGTCACTGCT 1140
 DB 105870 CCCGCACTGACCAAAATCTTAAATAGACTCAAAATCTCTTGTGTTAGTCACTGCT 105811

QY 1141 AGGAAAGTTGGGAGAGATAGTATGGAAGAAAGGCACTCTCGTTAGTCTTCC 1200
 DB 105810 AGGAAAGTTGGGAGAGATAGTATGGAAGAAAGGCACTCTCGTTAGTCTTCC 105751

QY 1201 GAAAGATTTACCATCTGAAACTGTTAAATCTTGAAGCAAGTATCAAAACAAGAG 1260
 DB 105750 GAAAGATTTACCATCTGAAACTGTTAAATCTTGAAGCAAGTATCAAAACAAGAG 105691

QY 1261 TGTTCACACACTTTACTGCTTAAACAAAAAATGTCCTCTGTCGACCAAGAAATTTTA 1320
 DB 105690 TGTTCACACACTTTACTGCTTAAACAAAAAATGTCCTCTGTCGACCAAGAAATTTTA 105631

QY 1321 TGATTAAGATTAATCTGTTAATCTGAGGCTATTAAGCTTGTGTAATAAGGCTG 1380
 DB 105630 TGATTAAGATTAATCTGTTAATCTGAGGCTATTAAGCTTGTGTAATAAGGCTG 105571

QY 1381 TAATCTGATTTCCAGCCTTAGACAAATTTAAGACGCTGAATGATGATGACCTAA 1440
 DB 105570 TAATCTGATTTCCAGCCTTAGACAAATTTAAGACGCTGAATGATGATGACCTAA 105511

QY 1441 TAAAGAAAAATTTGGATGATTTATGTCATTTCTAGACCAATTAACCATCCAGAGG 1500
 DB 105510 TAAAGAAAAATTTGGATGATTTATGTCATTTCTAGACCAATTAACCATCCAGAGG 105451

QY 1501 ACTTGGCAAAACCAATTTCTCAATTTGAGTATCTGAAGCAAGTTGCTTATGCTCAAT 1560
 DB 105450 ACTTGGCAAAACCAATTTCTCAATTTGAGTATCTGAAGCAAGTTGCTTATGCTCAAT 105391

QY 1561 ACCTGATTAAGTAAACAGCTGATGTTTGAATTTGATGAAATGATATATCACTGA 1620
 DB 105390 ACCTGATTAAGTAAACAGCTGATGTTTGAATTTGATGAAATGATATATCACTGA 105331

QY 1621 TGAAGAGATGCATATGTAACGCTCATATGAGCCATAGTCACTGATTTGAAAAAGATAG 1680
 DB 105330 TGAAGAGATGCATATGTAACGCTCATATGAGCCATAGTCACTGATTTGAAAAAGATAG 105271

QY 1681 CCTTCTGATTAAGAAAAAGTTGCAAGCTCAAGCTTATTAAGAAAAAGGATTCCTTAC 1740
 DB 105270 CCTTCTGATTAAGAAAAAGTTGCAAGCTCAAGCTTATTAAGAAAAAGGATTCCTTAC 105211

QY 1741 TCCATCTCCAGACGAGATGTTAAAGCAAACTGGAAGTATGAGCAGCAGCTATTTA 1800
 DB 105210 TCCATCTCCAGACGAGATGTTAAAGCAAACTGGAAGTATGAGCAGCAGCTATTTA 105151

QY 1801 CAATCGTGTGAAGGGGAAAAAGCAATTCACATGCTTGCATTCATATATGTTGAGCA 1860
 DB 105150 CAATCGTGTGAAGGGGAAAAAGCAATTCACATGCTTGCATTCATATATGTTGAGCA 105091

QY 1861 TACAGTTGAGTTAAAAACGGTAAATTTGATTATTCCTCATTAAGGATCATTAACAATAT 1920
 DB 105090 TACAGTTGAGTTAAAAACGGTAAATTTGATTATTCCTCATTAAGGATCATTAACAATAT 105031

QY 1921 TAAATTTGCTTGTGATGATCACAATCAAACTCCAAATGCTTATACCTTGAAGA 1980
 DB 105030 TAAATTTGCTTGTGATGATCACAATCAAACTCCAAATGCTTATACCTTGAAGA 104971

QY 1981 TTTGTTTGGAGAGATTAAGTACTAGTAACACCTTGACGAAGCTTCCATTTCAATGA 2040
 DB 104970 TTTGTTTGGAGAGATTAAGTACTAGTAACACCTTGACGAAGCTTCCATTTCAATGA 104911

OY	2041	TTGATGGGCAATGCCAGTACATGTGTTAGCAAGAAAGACACAGTAGAATCCAAA	2100	CDS	complement (91. .738)
					/gene="spr1056"
					/codon_start=1
Db	104910	TGGATGGGCAATGCCAGTACATGTGTTAGCAAGAAAGACACAGTAGAATCCAAA	104851		/transl_table=11
OY	2101	TAAAGACTTCAAAAGCGGATGAAGACCGATAGAGAAACACTGCTGAGCCAGAAAGTCCC	2160		/label="spr1056"
Db	104850	TAAAGACTTCAAAAGCGGATGAAGACCGATAGAGAAACACTGCTGAGCCAGAAAGTCCC	104791		/product="Hypothetical protein"
OY	2161	TCAAGTAGAGACTGAAAAAGTGAAGGCCCACTCAAAAGAGAGAAAGTTTGCTTGCGGA	2220		/protein_id="AAK9860.1"
Db	104790	TCAAGTAGAGACTGAAAAAGTGAAGGCCCACTCAAAAGAGAGAAAGTTTGCTTGCGGA	104731	gene	/db_xref="GI:15458678"
OY	2221	AGTAAAGGATTTCTAGCTGGAAGCCCAATGCAACAGAAACTTAGCTGGTTTACGAAATTA	2280		/translation="MTTKKANIITIEEYIEMSEVDVFNBAVNEFTSDTCOLANSYQSL KFPDKNFSGDLLFTWKSPLVEEGYIGRDSQVNLRAVIGNI FPNVLTNRKYSIN MNRGCMGDPHPDFDIYLDHAKYAAEOKVNNIKKEYPLKRALIHOENLYRFLFSN FDDPLEKXNYLKTIVQVSKETPFPSMDNMRKNISEKIIIFERGSMALDKSNYK"
Db	104730	AGTAAAGGATTTCTAGCTGGAAGCCCAATGCAACAGAAACTTAGCTGGTTTACGAAATTA	104671	CDS	complement (752. .1465)
OY	2281	TTTGACTCTTCAAAATTATGATTAACATATGATATGCGAAGAGCAAAAATTAATTCTGC	2340		/gene="spr1057"
Db	104670	TTTGACTCTTCAAAATTATGATTAACATATGATATGCGAAGAGCAAAAATTAATTCTGC	104611		/codon_start=1
OY	2341	GTTGTTAAAGAGATTAATCTTCACTCTGTAAGTAAAGAAAATTAAC	2389		/transl_table=11
Db	104610	GTTGTTAAAGAGATTAATCTTCACTCTGTAAGTAAAGAAAATTAAC	104562		/product="Conserved hypothetical protein"
RESULT 9	AE008479/c	10320 bp	DNA	linear	/protein_id="AAK9861.1"
LOCUS	AE008479	10320 bp	DNA	linear	/db_xref="GI:15458679"
DEFINITION	Streptococcus pneumoniae R6 section 95 of 184 of the complete genome.				/translation="MEYKRLFDLHDHTLDPDAEDVALTQLKEEGVADIQAKDYY VPMKALMKDLELKIKSKOELVNTRESRLSHRQEQGSFLAORVPIYLAQOCOTLS GAHDLDSLIERDLYAATNGITAIQTGRLAOSGLVFNQVPILEHLOTQKPDALF YKIGQOIAGFSKEXETLMIGDSLADIQGGNAGIDITIWPNPHELNHTQAOPYEVY SYODLDCDKNILEKITP"
ACCESSION	AE008479	AE007317			complement (1923. .2081)
VERSION	AE008479.1	GI:15458677			/gene="spr1058"
KEYWORDS					complement (1923. .2081)
SOURCE	Streptococcus pneumoniae R6				/gene="spr1059"
ORGANISM	Streptococcus pneumoniae R6				/codon_start=1
	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.				/transl_table=11
REFERENCE	1 (bases 1 to 10320)				/label="spr1058"
AUTHORS	Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszczak, L., Burgett, S., Dehoff, B.S., Estrem, S., Fritze, L., Fu, D.-J., Fuller, W., Gerlinger, C., Gilmore, R., Glass, J.S., Khoja, H., Kraft, A., Lagace, R., LeBlanc, D.J., Lee, L.N., Leikowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicot, T.I., Norris, F.H., O'Garra, M., Peery, R., Robertson, G.T., Rockey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L. and Glass, J.I.				/product="Hypothetical protein"
	Genome of the bacterium Streptococcus pneumoniae strain R6				/protein_id="AAK9863.1"
TITLE	J. Bacteriol. 183 (19), 5709-5717 (2001)				/db_xref="GI:15458681"
JOURNAL	21429245				/translation="IMESIGIGLVISHSKHIAGVVELISKVAKDVPIYVGTEGG GIGRSPD"
MEDLINE	1544234				complement (2151. .2306)
REFERENCE	2 (bases 1 to 10320)				/gene="spr1059"
AUTHORS	Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszczak, L., Burgett, S., Dehoff, B.S., Estrem, S., Fritze, L., Fu, D.-J., Fuller, W., Gerlinger, C., Gilmore, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., Lagace, R., LeBlanc, D.J., Lee, L.N., Leikowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicot, T.I., Norris, F.H., O'Garra, M., Peery, R., Robertson, G.T., Rockey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L. and Glass, J.I.				complement (2151. .2306)
	Genome of the bacterium Streptococcus pneumoniae strain R6				/gene="spr1059"
TITLE	J. Bacteriol. 183 (19), 5709-5717 (2001)				/codon_start=1
JOURNAL	21429245				/transl_table=11
MEDLINE	1544234				/label="spr1059"
REFERENCE	2 (bases 1 to 10320)				/product="Hypothetical protein"
AUTHORS	Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszczak, L., Burgett, S., Dehoff, B.S., Estrem, S., Fritze, L., Fu, D.-J., Fuller, W., Gerlinger, C., Gilmore, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., Lagace, R., LeBlanc, D.J., Lee, L.N., Leikowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicot, T.I., Norris, F.H., O'Garra, M., Peery, R., Robertson, G.T., Rockey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L. and Glass, J.I.				/protein_id="AAK9863.1"
	Genome of the bacterium Streptococcus pneumoniae strain R6				/db_xref="GI:15458681"
TITLE	J. Bacteriol. 183 (19), 5709-5717 (2001)				/translation="IMESIGIGLVISHSKHIAGVVELISKVAKDVPIYVGTEGG GIGRSPD"
JOURNAL	21429245				complement (2542. .2613)
MEDLINE	1544234				/rpt_family="RUP element"
REFERENCE	2 (bases 1 to 10320)				/rpt_type="dispersed"
AUTHORS	Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszczak, L., Burgett, S., Dehoff, B.S., Estrem, S., Fritze, L., Fu, D.-J., Fuller, W., Gerlinger, C., Gilmore, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., Lagace, R., LeBlanc, D.J., Lee, L.N., Leikowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicot, T.I., Norris, F.H., O'Garra, M., Peery, R., Robertson, G.T., Rockey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L. and Glass, J.I.				/rpt_unit="Rup_A"
	Genome of the bacterium Streptococcus pneumoniae strain R6				complement (2725. .5292)
TITLE	J. Bacteriol. 183 (19), 5709-5717 (2001)				/gene="phbA"
JOURNAL	21429245				/note="synonym: spr1060"
MEDLINE	1544234				complement (2725. .5292)
REFERENCE	2 (bases 1 to 10320)				/gene="phbA"
AUTHORS	Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszczak, L., Burgett, S., Dehoff, B.S., Estrem, S., Fritze, L., Fu, D.-J., Fuller, W., Gerlinger, C., Gilmore, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., Lagace, R., LeBlanc, D.J., Lee, L.N., Leikowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicot, T.I., Norris, F.H., O'Garra, M., Peery, R., Robertson, G.T., Rockey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L. and Glass, J.I.				/codon_start=1
	Genome of the bacterium Streptococcus pneumoniae strain R6				/transl_table=11
TITLE	J. Bacteriol. 183 (19), 5709-5717 (2001)				/label="spr1060"
JOURNAL	21429245				/product="Histidine Motif-Containing protein"
MEDLINE	1544234				/protein_id="AAK9864.1"
REFERENCE	2 (bases 1 to 10320)				/db_xref="GI:15458682"
AUTHORS	Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszczak, L., Burgett, S., Dehoff, B.S., Estrem, S., Fritze, L., Fu, D.-J., Fuller, W., Gerlinger, C., Gilmore, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., Lagace, R., LeBlanc, D.J., Lee, L.N., Leikowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicot, T.I., Norris, F.H., O'Garra, M., Peery, R., Robertson, G.T., Rockey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L. and Glass, J.I.				/translation="IMQIYLRKERMKINKYLAGSVATLVLSVCAVELGLHQATVK ENNRVYIDGKOATKXTENLTPDEVSREGINAEQIVKITDQGYVSHGHHYVYNG KVPYDALISELLMKDKNVOLKDEDDISEIKGGVIVADGGYVYVLDAAHADVTRK EAINRQKQHSQHREGCTPRNDAAVALARBOGRTTDDGYIFNASDIIEPTGDAVYIP HGDIHYIIPKNEUSASELAAKAPLSGRGNLSRTYTRRQNSDNTSTNNVPSVANG TTNTNTSNNSTNSQASNDIDSLIKLYKLPLSQHVSDDGLIFDPAQITSTYANG VAVPHGDHYHPIPYSQLSPLEKRLRIIPIRYSNMHWPPSRPEOPSPSPSPSPSP
	Genome of the bacterium Streptococcus pneumoniae strain R6				complement (2725. .5292)
TITLE	J. Bacteriol. 183 (19), 5709-5717 (2001)				/gene="phbA"
JOURNAL	21429245				/note="synonym: spr1060"
MEDLINE	1544234				complement (2725. .5292)
REFERENCE	2 (bases 1 to 10320)				/gene="phbA"
AUTHORS	Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszczak, L., Burgett, S., Dehoff, B.S., Estrem, S., Fritze, L., Fu, D.-J., Fuller, W., Gerlinger, C., Gilmore, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., Lagace, R., LeBlanc, D.J., Lee, L.N., Leikowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicot, T.I., Norris, F.H., O'Garra, M., Peery, R., Robertson, G.T., Rockey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L. and Glass, J.I.				/codon_start=1
	Genome of the bacterium Streptococcus pneumoniae strain R6				/transl_table=11
TITLE	J. Bacteriol. 183 (19), 5709-5717 (2001)				/label="spr1060"
JOURNAL	21429245				/product="Histidine Motif-Containing protein"
MEDLINE	1544234				/protein_id="AAK9864.1"
REFERENCE	2 (bases 1 to 10320)				/db_xref="GI:15458682"
AUTHORS	Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszczak, L., Burgett, S., Dehoff, B.S., Estrem, S., Fritze, L., Fu, D.-J., Fuller, W., Gerlinger, C., Gilmore, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., Lagace, R., LeBlanc, D.J., Lee, L.N., Leikowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicot, T.I., Norris, F.H., O'Garra, M., Peery, R., Robertson, G.T., Rockey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L. and Glass, J.I.				/translation="IMQIYLRKERMKINKYLAGSVATLVLSVCAVELGLHQATVK ENNRVYIDGKOATKXTENLTPDEVSREGINAEQIVKITDQGYVSHGHHYVYNG KVPYDALISELLMKDKNVOLKDEDDISEIKGGVIVADGGYVYVLDAAHADVTRK EAINRQKQHSQHREGCTPRNDAAVALARBOGRTTDDGYIFNASDIIEPTGDAVYIP HGDIHYIIPKNEUSASELAAKAPLSGRGNLSRTYTRRQNSDNTSTNNVPSVANG TTNTNTSNNSTNSQASNDIDSLIKLYKLPLSQHVSDDGLIFDPAQITSTYANG VAVPHGDHYHPIPYSQLSPLEKRLRIIPIRYSNMHWPPSRPEOPSPSPSPSPSP
	Genome of the bacterium Streptococcus pneumoniae strain R6				complement (2725. .5292)
TITLE	J. Bacteriol. 183 (19), 5709-5717 (2001)				/gene="phbA"
JOURNAL	21429245				/note="synonym: spr1060"
MEDLINE	1544234				complement (2725. .5292)
REFERENCE	2 (bases 1 to 10320)				/gene="phbA"
AUTHORS	Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszczak, L., Burgett, S., Dehoff, B.S., Estrem, S., Fritze, L., Fu, D.-J., Fuller, W., Gerlinger, C., Gilmore, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., Lagace, R., LeBlanc, D.J., Lee, L.N., Leikowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicot, T.I., Norris, F.H., O'Garra, M., Peery, R., Robertson, G.T., Rockey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L. and Glass, J.I.				/codon_start=1
	Genome of the bacterium Streptococcus pneumoniae strain R6				/transl_table=11
TITLE	J. Bacteriol. 183 (19), 5709-5717 (2001)				/label="spr1060"
JOURNAL	21429245				/product="Histidine Motif-Containing protein"
MEDLINE	1544234				/protein_id="AAK9864.1"
REFERENCE	2 (bases 1 to 10320)				/db_xref="GI:15458682"
AUTHORS	Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszczak, L., Burgett, S., Dehoff, B.S., Estrem, S., Fritze, L., Fu, D.-J., Fuller, W., Gerlinger, C., Gilmore, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., Lagace, R., LeBlanc, D.J., Lee, L.N., Leikowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicot, T.I., Norris, F.H., O'Garra, M., Peery, R., Robertson, G.T., Rockey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L. and Glass, J.I.				/translation="IMQIYLRKERMKINKYLAGSVATLVLSVCAVELGLHQATVK ENNRVYIDGKOATKXTENLTPDEVSREGINAEQIVKITDQGYVSHGHHYVYNG KVPYDALISELLMKDKNVOLKDEDDISEIKGGVIVADGGYVYVLDAAHADVTRK EAINRQKQHSQHREGCTPRNDAAVALARBOGRTTDDGYIFNASDIIEPTGDAVYIP HGDIHYIIPKNEUSASELAAKAPLSGRGNLSRTYTRRQNSDNTSTNNVPSVANG TTNTNTSNNSTNSQASNDIDSLIKLYKLPLSQHVSDDGLIFDPAQITSTYANG VAVPHGDHYHPIPYSQLSPLEKRLRIIPIRYSNMHWPPSRPEOPSPSPSPSPSP
	Genome of the bacterium Streptococcus pneumoniae strain R6				complement (2725. .5292)
TITLE	J. Bacteriol. 183 (19), 5709-5717 (2001)				/gene="phbA"
JOURNAL	21429245				/note="synonym: spr1060"
MEDLINE	1544234				complement (2725. .5292)
REFERENCE	2 (bases 1 to 10320)				/gene="phbA"
AUTHORS	Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszczak, L., Burgett, S., Dehoff, B.S., Estrem, S., Fritze, L., Fu, D.-J., Fuller, W., Gerlinger, C., Gilmore, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., Lagace, R., LeBlanc, D.J., Lee, L.N., Leikowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicot, T.I., Norris, F.H., O'Garra, M., Peery, R., Robertson, G.T., Rockey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L. and Glass, J.I.				/codon_start=1
	Genome of the bacterium Streptococcus pneumoniae strain R6				/transl_table=11
TITLE	J. Bacteriol. 183 (19), 5709-5717 (2001)				/label="spr1060"
JOURNAL	21429245				/product="Histidine Motif-Containing protein"
MEDLINE	1544234				/protein_id="AAK9864.1"
REFERENCE	2 (bases 1 to 10320)				/db_xref="GI:15458682"
AUTHORS	Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszczak, L., Burgett, S., Dehoff, B.S., Estrem, S., Fritze, L., Fu, D.-J., Fuller, W., Gerlinger, C., Gilmore, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., Lagace, R., LeBlanc, D.J., Lee, L.N., Leikowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicot, T.I., Norris, F.H., O'Garra, M., Peery, R., Robertson, G.T., Rockey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L. and Glass, J.I.				/translation="IMQIYLRKERMKINKYLAGSVATLVLSVCAVELGLHQATVK ENNRVYIDGKOATKXTENLTPDEVSREGINAEQIVKITDQGYVSHGHHYVYNG KVPYDALISELLMKDKNVOLKDEDDISEIKGGVIVADGGYVYVLDAAHADVTRK EAINRQKQHSQHREGCTPRNDAAVALARBOGRTTDDGYIFNASDIIEPTGDAVYIP HGDIHYIIPKNEUSASELAAKAPLSGRGNLSRTYTRRQNSDNTSTNNVPSVANG TTNTNTSNNSTNSQASNDIDSLIKLYKLPLSQHVSDDGLIFDPAQITSTYANG VAVPHGDHYHPIPYSQLSPLEKRLRIIPIRYSNMHWPPSRPEOPSPSPSPSPSP
	Genome of the bacterium Streptococcus pneumoniae strain R6				complement (2725. .5292)
TITLE	J. Bacteriol. 183 (19), 5709-5717 (2001)				/gene="phbA"
JOURNAL	21429245				/note="synonym: spr1060"
MEDLINE	1544234				complement (2725. .5292)
REFERENCE	2 (bases 1 to 10320)				/gene="phbA"
AUTHORS	Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszczak, L., Burgett, S., Dehoff, B.S., Estrem, S., Fritze, L., Fu, D.-J., Fuller, W., Gerlinger, C., Gilmore, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., Lagace, R., LeBlanc, D.J., Lee, L.N., Leikowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicot, T.I., Norris, F.H., O'Garra, M., Peery, R., Robertson, G.T., Rockey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L. and Glass, J.I.				/codon_start=1
	Genome of the bacterium Streptococcus pneumoniae strain R6				/transl_table=11
TITLE	J. Bacteriol. 183 (19), 5709-5717 (2001)				/label="spr1060"
JOURNAL	21429245				/product="Histidine Motif-Containing protein"
MEDLINE	1544234				/protein_id="AAK9864.1"
REFERENCE	2 (bases 1 to 10320)				/db_xref="GI:15458682"
AUTHORS	Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszczak, L., Burgett, S., Dehoff, B.S., Estrem, S., Fritze, L., Fu, D.-J., Fuller, W., Gerlinger, C., Gilmore, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., Lagace, R., LeBlanc, D.J., Lee, L.N., Leikowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicot, T.I., Norris, F.H., O'Garra, M., Peery, R., Robertson, G.T., Rockey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L. and Glass, J.I.				/translation="IMQIYLRKERMKINKYLAGSVATLVLSVCAVELGLHQATVK ENNRVYIDGKOATKXTENLTPDEVSREGINAEQIVKITDQGYVSHGHHYVYNG KVPYDALISELLMKDKNVOLKDEDDISEIKGGVIVADGGYVYVLDAAHADVTRK EAINRQKQHSQHREGCTPRNDAAVALARBOGRTTDDGYIFNASDIIEPTGDAVYIP HGDIHYIIPKNEUSASELAAKAPLSGRGNLSRTYTRRQNSDNTSTNNVPSVANG TTNTNTSNNSTNSQASNDIDSLIKLYKLPLSQHVSDDGLIFDPAQITSTYANG VAVPHGDHYHPIPYSQLSPLEKRLRIIPIRYSNMHWPPSRPEOPSPSPSPSPSP
	Genome of the bacterium Streptococcus pneumoniae strain R6				complement (2725. .5292)
TITLE	J. Bacteriol. 183 (19), 5709-5717 (2001)				/gene="phbA"
JOURNAL	21429245				/note="synonym: spr1060"
MEDLINE	1544234				complement (2725. .5292)
REFERENCE	2 (bases 1 to 10320)				/gene="phbA"
AUTHORS	Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszczak, L., Burgett, S., Dehoff, B.S., Estrem, S., Fritze, L., Fu, D.-J., Fuller, W., Gerlinger, C., Gilmore, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., Lagace, R., LeBlanc, D.J., Lee, L.N., Leikowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicot, T.I., Norris, F.H., O'Garra, M., Peery, R., Robertson, G.T., Rockey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L. and Glass, J.I.				/codon_start=1
	Genome of the bacterium Streptococcus pneumoniae strain R6				/transl_table=11
TITLE	J. Bacteriol. 183 (19), 5709-5717 (2001)				/label="spr1060"
JOURNAL	21429245				/product="Histidine Motif-Containing protein"
MEDLINE	1544234				/protein_id="AAK9864.1"
REFERENCE	2 (bases 1 to 10320)				/db_xref="GI:15458682"
AUTHORS	Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszczak, L., Burgett, S., Dehoff, B.S., Estrem, S., Fritze, L., Fu, D.-J., Fuller, W., Gerlinger, C., Gilmore, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., Lagace, R., LeBlanc, D.J., Lee, L.N., Leikowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicot, T.I., Norris, F.H., O'Garra, M., Peery, R., Robertson, G.T., Rockey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Baltz, R.H., Jaskunas, S.R., Rosteck, P.R. Jr., Skatrud, P.L. and Glass, J.I.				/translation="IMQIYLRKERMKINKYLAGSVATLVLSVCAVELGLHQATVK ENNRVYIDGKOATKXTENLTPDEVSREGINAEQIVKITDQGYVSHGHHYVYNG KVPYDALISELLMKDKNVOLKDEDDISEIKGGVIVADGGYVYVLDAAHADVTRK EAINRQKQHSQHREGCTPRNDAAVALARBOGRTTDDGYIFNASDIIEPTGDAVYIP HGDIHYIIPKNEUSASELAAKAPLSGRGNLSRTYTRRQNSDNTSTNNVPSVANG TTNTNTSNNSTNSQASNDIDSLIKLYKLPLSQHVSDDGLIFDPAQITSTYANG VAVPHGDHYHPIPYSQLSPLEKRLRIIPIRYSNMHWPPSRPEOPSPSPSPSPSP
	Genome of the bacterium Streptococcus pneumoniae strain R6				complement (2725. .5292)
TITLE	J. Bacteriol. 183 (19), 5709-5717 (2001)				/gene="phbA"
JOURNAL	21429245				/note="synonym: spr1060"
MEDLINE	1544234				complement (2725. .5292)
REFERENCE	2 (bases 1 to 10320)				/gene="phbA"
AUTHORS	Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszczak, L., Burgett, S., Dehoff, B.S., Estrem, S., Fritze, L., Fu, D.-J., Fuller, W., Gerlinger, C., Gilmore, R., Glass, J.S., Hann, A., Khoja, H., Kraft, A., Lagace, R., LeBlanc, D.J., Lee, L.N., Leikowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Nicot, T.I., Norris, F.H., O'Garra, M., Peery, R., Robertson, G.T., Rockey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G.,				

QPAENPQAPSPNIDEKLVKEAVRKVGQGYFEENGVPRIYIPAKDLSAETAJDIDSKL
AKQESLSHKLAKKIDTLPSSDREFNKAYDLARJHODLMDKRGQVDFEALDNLIER
LKDVSSDRYKLVDDILAFAPLRHPERLQKPAQATYDDDEIQVAKLAKYTGEGYI
FDPDITSDDEGAAYTPHMTSHHWIKDLSAEBAQAQAKKEKGLPSPDHDDSG
NTEAKGAIYARVKAARKVPLDPMYNOQVYENKSLIIPHYDHNKIFTEFDE
GLYEPKYSLEDLATAKYIYEHNERPHSNGNAGNSDHPKQKQADQNOUEKP
NEEKQTEKPEEETREEKPOSEKPEPPEEPEESPEESPESEEPVTEYVKE
KLREADDLGKQNPDIKSNAKETLGLKNLLFGTDNNITMAEAKLLALKESK"
complement (5417. .7903)
/gene="phla"
/note="synonym: spr1061"
complement (5417. .7903)
/gene="phla"
/codon_start=1
/transl_table=11
/label=spr1061
/product="Pneumococcal histidine triad protein A precursor"

/protein_id="AAK9865.1"
/db_xref="GI:15458683"
/translation="MOLEISNRKRVSMKINKKYLVGSAALILSVCSYELGYOARTY
KENNVSYIDGKOAKTENTLPDEVSREGINABOIVYKITDQCVYVSHQHGAYYV
GKVPDAIFSEELMKDPRYKIKDEIDYNEVKGIVYKIDGKTYIYLDADAHADVRT
KEIIRQKQESHQREGGTPRNDGVALARSQGRYTTDDGYIENASDIIEDTGAYIV
PHGDHYHYPKNEISASELAEAFLSGRGNLSNRTRYRONSNTSTNWPVSYP
GTTNTNTSNNTNSQASQNDISLKKOLYKLPLOSHESEDGLVPDPAOTRSTR
GAVAPHGHHYHIFYSOMSELEERLARIPLKRSNMHPDPRPPOPSPOPPSPG
POPAPLKIDNSLSVQLVRKVGQYVEEKGISRYVAPADLPSEYTKNLESKSKO
ESVSHITLAKKENVAPRODERFDKAYNLITAEKALFEKGNDSFOALDKLERLND
ESTNKEKLVDDILAFAPITPERLGRNSQIEYEDERLAQLADKXTYISDGYIFDE
HDIIISDEDAVYTPMGHSHWIKDLSDEKVAQAAYTKEGIIIPSPDPAKANTP
GDSAAIYARVYGEKRIPLVLPYVHEHVEYKGNLLIPKDHYNIKFAVFDHTY
KAPNGYLTLEDIYATIKYVHEHDERPHSNDGNGASBHVGLKXHSDDPNKFKADE
PVEETPAPEVPOVTEKVEAQLKEAVULAVYVDSLKANAETLGLRNLTLQIM
DNMSIMAEAKLLALKSNBSVSKEKIN"
complement (8122. .9855)
/gene="ptst1"
/note="synonym: spr1062"
complement (8122. .9855)
/gene="ptst1"
/BC_number="2.7.3.9"
/codon_start=1
/transl_table=11
/label=spr1062
/product="PTS enzyme I"
/protein_id="AAK9866.1"
/db_xref="GI:15458684"
/translation="MTEMLKGIASDGVAVAKAYLLVOPDLSFETITVEDTNAEARL
DAALQASODELSVIREKAVGLTGEBAQVFDHMLVADPEMIISOIKETIRAKRYNAE
AGLKEVTDMFTIPEGMEDNPYMOERADIRVTRKVLANLGLKPNPASINEVIV
IAHDLTPSDTQOLDKNFVAKAPVNTICGRTSHSAIMARPLEIAVYGTNNITVEVGD
ILAVNGITGEVITINPTDEQABFKAGERYAKOKABMLLKQAGTVITADGKFEILAN
IGTPDVEGVNNGAEAVGLRTEFLYVDSODFPEDDEYERKAVLSEGMKGPVVR
ITMDIGDKELPFYDMHBMNPLGFRLRISISETGDMFROIALLRASVHGQLRI
MFPMVALKEFERAAKAVDEEKANLLAGVAVADNIQVIMIEIPAAALMDQFKEV
DEFSIGTNDLIQYTAADRNMNOVLYLOPVNSILIRLNNVYIKAAHGEKAGQCE
MAGDOAVPLVGMGLDERSMATSVLTRSLMKLDLPAKMEYANRALTEGSTIEEV
LELQKEYVNFV"
complement (9861. .10124)
/gene="ptstH"
/note="synonym: spr1063"
complement (9861. .10124)
/gene="ptstH"
/codon_start=1
/transl_table=11
/label=spr1063
/product="Histidine-containing phosphocARRIER protein of
the PTS"
/protein_id="AAK9867.1"
/db_xref="GI:15458685"
/translation="WASKDFHVAETGIHARPATLLVOTASKFASDITLEYKGSVNL
KSIMGMSLIGVGGADVTISAGADADDAIAAISETMEKEGIA"

BASE COUNT

2851 a 2091 c 1862 g 3516 t

ORIGIN

Query Match 99.0%; Score 2365.6; DB 1; Length 10320;

Best Local Similarity 99.4%; Pred. No. 0; Mismatches 15; Indels 0; Gaps 0;

Matches 2374; Conservative 0; Nucleotides 15; Indels 0; Gaps 0;

QY	1	TTCTTACGAGTTGGACCTGTATCAAGCTAAGACGTTAAGAAATAATCGTGTTCCTA	60
DB	7808	TTCTTACGAGTTGGACCTGTATCAAGCTAAGACGTTAAGAAATAATCGTGTTCCTA	7749
QY	61	TATACATGAAAAACAGCGAGCAAAAAACGAGATTTCCTCTGATGAGTTAGCA	120
DB	7748	TATACATGAAAAACAGCGAGCAAAAAACGAGATTTCCTCTGATGAGTTAGCA	7689
QY	121	GGTGAAGAAATCAATGCTGAGCAAAATGCTATCAAGTAACAGCCAAAGCTATGTCAC	180
DB	7688	GGTGAAGAAATCAATGCTGAGCAAAATGCTATCAAGTAACAGCCAAAGCTATGTCAC	7629
QY	181	TTACATGAGCGACCACTATCATTTATTAATGTAAGGTTCCCTTATGACGCTATCATGAC	240
DB	7628	TTACATGAGCGACCACTATCATTTATTAATGTAAGGTTCCCTTATGATGCGATTTTCAG	7569
QY	241	TGAAGATTACTCATGAAGAATCCAACTATPAGCTAAAGATGAGATATTGTTAATGA	300
DB	7568	TGAAGATTACTCATGAAGAATCCAACTATPAGCTAAAGATGAGATATTGTTAATGA	7509
QY	301	GGTCAAGGCTGATATGTTATCAAGTATGAGTAATAATATATCTTTACCTTAAGATGC	360
DB	7508	GGTCAAGGCTGATATGTTATCAAGTATGAGTAATAATATATCTTTACCTTAAGATGC	7449
QY	361	TGCCACCCGGAATAACGTCCTCAACAAAGAGAAATCAATGACCAAAAAACAGACATAG	420
DB	7448	TGCCACCCGGAATAACGTCCTCAACAAAGAGAAATCAATGACCAAAAAACAGACATAG	7389
QY	421	TCAACATGCTGAGAGTGAATCTCCAAAGAACGATGCTGCTGCTGCGACGTTCCGA	480
DB	7388	TCAACATGCTGAGAGTGAATCTCCAAAGAACGATGCTGCTGCTGCGACGTTCCGA	7329
QY	481	AGGAGCTATATCTACAGATGATGTTATATCTTTAATGCTTCTGATATCATGAGATATC	540
DB	7328	AGGAGCTATATCTACAGATGATGTTATATCTTTAATGCTTCTGATATCATGAGATATC	7269
QY	541	TGCTATGCTTATATATCGTTCCTCATGAGATCATTAACATTACCTCTAAGATGATG	600
DB	7268	TGCTATGCTTATATATCGTTCCTCATGAGATCATTAACATTACCTCTAAGATGATG	7209
QY	601	ATCAGCTGAGGAGTGGCTGTCAGAAAGCCTTCTCATCTGCTGAGGAATCTGTCAA	660
DB	7208	ATCAGCTGAGGAGTGGCTGTCAGAAAGCCTTCTCATCTGCTGAGGAATCTGTCAA	7149
QY	661	TTCAAGAACTTATGCGCCGACAAATAGGATATACCTTCAAGAACTGGGTACCTTC	720
DB	7148	TTCAAGAACTTATGCGCCGACAAATAGGATATACCTTCAAGAACTGGGTACCTTC	7089
QY	721	TGTAAGCAATCAGGAATCAAAATACATCAACAAAGCAACAAACAGCACTTAAGTCA	780
DB	7088	TGTAAGCAATCAGGAATCAAAATACATCAACAAAGCAACAAACAGCACTTAAGTCA	7029
QY	781	AGCAAGTCAAAAGTAATGACATGATATGCTCTTGAAGACGCTTCAAACTGCTTGAAG	840
DB	7028	AGCAAGTCAAAAGTAATGACATGATATGCTCTTGAAGACGCTTCAAACTGCTTGAAG	6969
QY	841	TCAAGCAATGATGAGATCTGATGCTGCTCTTATATCAGCAAAATCAAGTCGAC	900
DB	6968	TCAAGCAATGATGAGATCTGATGCTGCTCTTATATCAGCAAAATCAAGTCGAC	6909
QY	901	AGCTAGAGTGTGAGTGCACACAGAGATCATTAACCTTCAATCCCTTACTCTCAAT	960
DB	6908	AGCTAGAGTGTGAGTGCACACAGAGATCATTAACCTTCAATCCCTTACTCTCAAT	6849
QY	961	GTCGTAATTGGAAGAAAGCAATCGCTGATATATCCCTTCTGTTATCGTTCAACCATG	1020
DB	6848	GTCGTAATTGGAAGAAAGCAATCGCTGATATATCCCTTCTGTTATCGTTCAACCATG	6789

```

OY 1021 GGTACAGATTCAAGGCCAGAACCAAGTCCACACGAGTCCGGAACCTAGTCCAGG 1080
DB 6788 GGTACAGATTCAAGGCCAGAACCAAGTCCACACGAGTCCGGAACCTAGTCCAGG 6729
OY 1081 CCCGCAACCTCGACCAAACTCTTAAATAGACTCAAAATCTTCTTGGTTAGTCACTGGT 1140
DB 6728 CCCGCAACCTCGACCAAACTCTTAAATAGACTCAAAATCTTCTTGGTTAGTCACTGGT 6669
OY 1141 ACGAAATTTGGGGAAGATATGTATTGGAAGAAAGGGCATCTCTGTTATGTTGTC 1200
DB 6668 ACGAAATTTGGGGAAGATATGTATTGGAAGAAAGGGCATCTCTGTTATGTTGTC 6609
OY 1201 GAAAGATTTCATCTGAAACCTGTTAAATCTTGAAGCAAGTATCAAAACAAGAGAG 1260
DB 6608 GAAAGATTTCATCTGAAACCTGTTAAATCTTGAAGCAAGTATCAAAACAAGAGAG 6549
OY 1261 TGTTCACACACTTTTAACTGCTAAAAAAGAAATGTTCTCTCTGTCGACCAAGATTTTA 1320
DB 6548 TGTTCACACACTTTTAACTGCTAAAAAAGAAATGTTCTCTCTGTCGACCAAGATTTTA 6489
OY 1321 TGATTAAGCATATATCTGTTTACTGAGGCTCATAAAGCTTGTGTAATTAAGGTCG 1380
DB 6488 TGATTAAGCATATATCTGTTTACTGAGGCTCATAAAGCTTGTGTAATTAAGGTCG 6429
OY 1381 TAATTCGATTCCAAAGCCTTAGACAAATATTAGACGCTTGAATGATGATGACTAA 1440
DB 6428 TAATTCGATTCCAAAGCCTTAGACAAATATTAGACGCTTGAATGATGATGACTAA 6369
OY 1441 TAAAGAAAAATTTGTAGATGATTTATTTGGCATCTTAGACCAATTAACCATCCAGAGCG 1500
DB 6368 TAAAGAAAAATTTGTAGATGATTTATTTGGCATCTTAGACCAATTAACCATCCAGAGCG 6309
OY 1501 ACTTGGCAAAACCAATTTCTCAATTTGAGTATTTCTGAACCAAGTTCGTTTCTCATTT 1560
DB 6308 ACTTGGCAAGCAAAATTTCTCAATTTGAGTATTTCTGAACCAAGTTCGTTTCTCATTT 6249
OY 1561 AGCTGATAGATATACACGTCAGATGTTTACATTTTGTATGAAACATGATATATAGTGA 1620
DB 6248 AGCTGATAGATATACACGTCAGATGTTTACATTTTGTATGAAACATGATATATAGTGA 6189
OY 1621 TGAAGAGATGATATGTAAAGCCTCATATAGGCAATAGTCACTGATTTGAAAAAGATAG 1680
DB 6188 TGAAGAGATGATATGTAAAGCCTCATATAGGCAATAGTCACTGATTTGAAAAAGATAG 6129
OY 1681 CTTTCTGATATAGAAAAAGTTGCGCTCAAGCCTTACTTAAAGAAAAAGTATCTTACC 1740
DB 6128 CTTTCTGATATAGAAAAAGTTGCGCTCAAGCCTTACTTAAAGAAAAAGTATCTTACC 6069
OY 1741 TCCATCTCGACGAGATGTTAAAGCAATTCGAAGTGAAGTATGAGCAGCTATTTA 1800
DB 6068 TCCATCTCGACGAGATGTTAAAGCAATTCGAAGTGAAGTATGAGCAGCTATTTA 6009
OY 1801 CAATCTGTGAAGGGGAAAAAGCAATTCGCTGCTGCTTCCATATATGTTGAGCA 1860
DB 6008 CAATCTGTGAAGGGGAAAAAGCAATTCGCTGCTGCTTCCATATATGTTGAGCA 5949
OY 1861 TACAGTTGAGTTAAAAAGGTAATTTGATTTTCTCTATTAAGATCTTACATATAT 1920
DB 5948 TACAGTTGAGTTAAAAAGGTAATTTGATTTTCTCTATTAAGATCTTACATATAT 5889
OY 1921 TAAATTTGCTGTTGATGATCACAATCAAAAGCTCAAAATGCTTACCTTGAAGA 1980
DB 5888 TAAATTTGCTGTTGATGATCACAATCAAAAGCTCAAAATGCTTACCTTGAAGA 5829
OY 1981 TTTGTTTGCAGATTAAGTACTAGTAGAACACCTCGTAGCAAGCTGCACATTTCAATGA 2040
DB 5828 TTTGTTTGCAGATTAAGTACTAGTAGAACACCTCGTAGCAAGCTGCACATTTCAATGA 5769
OY 2041 TGGATGGGCAATGCGCAGTGAAGATGTTAGGCAAGAAAGCACAAGTGAAGATCCAA 2100
DB 5768 TGGATGGGCAATGCGCAGTGAAGATGTTAGGCAAGAAAGCACAAGTGAAGATCCAA 5709

```

```

OY 2101 TAAGACTTCAAAAGCGATGAGAGAGCCAGTAGAGAAAAACCTGCTAGCCAGAGTCCC 2160
DB 5708 TAAGACTTCAAAAGCGATGAGAGAGCCAGTAGAGAAAAACCTGCTAGCCAGAGTCCC 5649
OY 2161 TCAAGTAGAGCTGAAAAAGTAGAACCCCACTCAAGAAAGCAGAAAGTTTGTGGCA 2220
DB 5648 TCAAGTAGAGCTGAAAAAGTAGAACCCCACTCAAGAAAGCAGAAAGTTTGTGGCA 5589
OY 2221 AGTAAGGATTTCTAGTCTGAAGCCATGCAACGAAACTCTAGCTGTTACGAATTA 2280
DB 5588 AGTAAGGATTTCTAGTCTGAAGCCATGCAACGAAACTCTAGCTGTTACGAATTA 5529
OY 2281 TTTGACTCTTCAATTTATGATTAACAATAGTATGATGAGCAGAAAGCAAAATTA 2340
DB 5528 TTTGACTCTTCAATTTATGATTAACAATAGTATGATGAGCAGAAAGCAAAATTA 5469
OY 2341 GTTGTAAAAAGAGTAAATCTTCTATCTGTAAGTAAAGAAAAATTAAC 2389
DB 5468 GTTGTAAAAAGAGTAAATCTTCTATCTGTAAGTAAAGAAAAATTAAC 5420

```

```

RESULT 10
SPNEU1915
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

```

```

REFERENCE
1
Dopazo, J., Mendoza, A., Herrero, J., Calafra, F., Humbert, Y.,
Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de
Francisco, M., Polissi, A., Buell, G., Feger, G., Garcia, E., Peitsch, M.,
and Garcia-Bustos, J. F.
Annotated draft genomic sequence from a Streptococcus pneumoniae
type 19F clinical isolate
Microb. Drug Resist. 7 (2), 99-125 (2001)

```

```

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

```

```

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

```

```

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

```

```

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

```

```

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

```

```

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

```

```

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

```

```

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

```

```

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

```

```

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

```

Db 17130 TATAGATGAAAAACAAGCAGCAGCAAAAAACGAGAAATTTGACTCCGATGAGAGTTAGCA 17189
 Qy 121 GGGTAAAGAAATCAATGCTGAGCAAAATGCTCATCAAGATTAACAGCCAGCTATGTCAC 180
 Db 17190 GGGTAAAGAAATCAATGCTGAGCAAAATGCTCATCAAGATTAACAGCCAGCTATGTCAC 17249
 Qy 181 TTCACATGGCGACCACTATCATTTATTAACATGTAAGGTTCCCTTATGACGCTATCATGAC 240
 Db 17250 TTCACATGGCGACCACTATCATTTATTAACATGTAAGGTTCCCTTATGAGCCGATTTTTCG 17309
 Qy 241 TGAAGAAATTTACTCATGAAAAGATCCAAACTATAAGCTAAAAAGATGAGATATTGTTAATGA 300
 Db 17310 TGAAGAAATTTACTCATGAAAAGATCCAAACTATAAGCTAAAAAGATGAGATATTGTTAATGA 17369
 Qy 301 GGTCAAGGTGATGATGTTATGTAAGTGAATGGAATAATCTATGTTTACTTAAGAGATG 360
 Db 17370 GGTCAAGGTGATGATGTTATGTAAGTGAATGGAATAATCTATGTTTACTTAAGAGATG 17429
 Qy 361 TGCCACGCGGATTAAGTCCGTACAAAAGAGAAATCAATCGACAAAACAAAGAGCATAG 420
 Db 17430 TGCCACGCGGATTAAGTCCGTACAAAAGAGAAATCAATCGACAAAACAAAGAGCATAG 17489
 Qy 421 TCAACATCGTGAAGGTGAAGCTCCAAAGAAACATGCTGCTGCTTGCCTTGGCAGCTTCGCA 480
 Db 17490 TCAACATCGTGAAGGTGAAGCTCCAAAGAAACATGCTGCTGCTTGCCTTGGCAGCTTCGCA 17549
 Qy 481 AGGAGCGTCTACTACAGATGATGCTTATATCTTTAATGCTTGTGATATCATAGAGATG 540
 Db 17550 AGGAGCGTCTACTACAGATGATGCTTATATCTTTAATGCTTGTGATATCATAGAGATG 17609
 Qy 541 TGGTATGCTTATATGCTTCTCATGAGATCATTTCCATTAATCTTCTTAAGATGAGTT 600
 Db 17610 TGGTATGCTTATATGCTTCTCATGAGATCATTTCCATTAATCTTCTTAAGATGAGTT 17669
 Qy 601 ATCAGTACGAGTGTGCTGCTGCTGCAAGAGCTTTCCTATGCTGCTGCAAGAAATCTGTCAAA 660
 Db 17670 ATCAGTACGAGTGTGCTGCTGCTGCAAGAGCTTTCCTATGCTGCTGCAAGAAATCTGTCAAA 17729
 Qy 661 TTCAGAACTTATGCTGCTGCTGCAAGAAATGCGATTAACCTTTAAGAACTGCGTACCTTC 720
 Db 17730 TTCAGAACTTATGCTGCTGCTGCAAGAAATGCGATTAACCTTTAAGAACTGCGTACCTTC 17789
 Qy 721 TGTAAAGCATCCAGAACTTAACAATTAACAACAAGCAACAGAACTTAACAAGTCA 780
 Db 17790 TGTAAAGCATCCAGAACTTAACAATTAACAACAAGCAACAGAACTTAACAAGTCA 17849
 Qy 781 AGCAAGTCAAAAGTAAATGACATTTGATGATGCTTGTGAAGACGCTCTCAAACTGCTTTGAG 840
 Db 17850 ARCAAGTCAAAAGTAAATGACATTTGATGATGCTTGTGAAGACGCTCTCAAACTGCTTTGAG 17909
 Qy 841 TCAAGCAGATGTAAGATCTGATGCTTGTCTTTGATCCAGCACAAAATCAAGTCAAGTCAAC 900
 Db 17910 TCAAGCAGATGTAAGATCTGATGCTTGTCTTTGATCCAGCACAAAATCAAGTCAAGTCAAC 17969
 Qy 901 AGCTAGAGTGTGAGTGCAGCAGAGATGATTAACAATCACTCAATCCCTTACTCAAT 960
 Db 17970 AGCTAGAGTGTGAGTGCAGCAGAGATGATTAACAATCACTCAATCCCTTACTCAAT 18029
 Qy 961 GTCTGAATTTGAAGAACGATGCTGATATTTCCCTTGTGTTATGCTTCAAACTGCTTCAAT 1020
 Db 18030 GTCTGAATTTGAAGAACGATGCTGATATTTCCCTTGTGTTATGCTTCAAACTGCTTCAAT 18089
 Qy 1021 GGTACCAAGTTCAAGGCGAAGCAACCAAGTCCACACGAGCTCCGGAACCTAGTCCAG 1080
 Db 18090 GGTACCAAGTTCAAGGCGAAGCAACCAAGTCCACACGAGCTCCGGAACCTAGTCCAG 18149
 Qy 1081 CCCGCACTGCAACCAATCTTAATAATAGCTCAATTTCTTGTGTTAGTCAAGTGT 1140
 Db 18150 CCCGCACTGCAACCAATCTTAATAATAGCTCAATTTCTTGTGTTAGTCAAGTGT 18209
 Qy 1141 ACGAAAGTTGGGAGAGATATGATTCGAGAAAAAGGAGCTCTGCTTATGCTTTG 1200

Db 18210 ACGAAAGTTGGGAGAGATATGATTCGAGAAAAAGGAGCACTCTGTTATGCTTTG 18269
 Qy 1201 GAAGATTTTACCATGTGAACCTGTTAAAAATCTTGAAGCAAGTATCAAAACAAGAG 1260
 Db 18270 GAAGATTTTACCATGTGAACCTGTTAAAAATCTTGAAGCAAGTATCAAAACAAGAG 18329
 Qy 1261 TGTTCACACACTTTTAACTGCTAAAAAGAAATGTTCTCCTGTCGACCAAGATTTTA 1320
 Db 18330 TGTTCACACACTTTTAACTGCTAAAAAGAAATGTTCTCCTGTCGACCAAGATTTTW 18389
 Qy 1321 TGATTAAGCATTAATCTGTTTACTGAGGCTCATTAAGCCTTTGTTGNAATTAAGGTCG 1380
 Db 18390 TGATTAAGCATTAATCTGTTTACTGAGGCTCATTAAGCCTTTGTTGNAATTAAGGTCG 18449
 Qy 1391 TAAATCTGATTTCCAGCCTTAAGAAATTAATTAAGCCTTGAATGATGATGAC -TA 1439
 Db 18450 TAAATCTGATTTCCAGCCTTGAAGAAATTTTAAGCCTTGAATGATGATGACCTTA 18509
 Qy 1440 ATTAAGAAAAATTTGATGATGTTTATGAGCTTCTAGCACCAATTAACCATCCAGAG 1499
 Db 18510 ATTAAGAAAAATTTGATGATGTTTATGAGCTTCTAGCACCAATTAACCATCCAGAG 18569
 Qy 1500 GACTTGGCAACCAAAATTTCAAAATTTGATGATGATTAAGAGCAAGCTGATTTGCTCAAT 1559
 Db 18570 GACTTGGCAACCAAAATTTCAAAATTTGATGATGATTAAGAGCAAGCTGATTTGCTCAAT 18629
 Qy 1560 TAGCTGATTAAGTATCAACGTCAGATGCTTACATTTTGTGATGAATGATTAATCAAGT 1619
 Db 18630 TAGCTGATTAAGTATCAACGTCAGATGCTTACATTTTGTGATGAATGATTAATCAAGT 18689
 Qy 1620 ATGAAGAGATGATATGATTAAGCCTCATTAAGGCGATGATCACTGATTTGAAAGATA 1679
 Db 18690 ATGAAGAGATGATATGATTAAGCCTCATTAAGGCGATGATCACTGATTTGAAAGATA 18749
 Qy 1680 GCTTTTGTGATGAAGAAAAAGTTGCAAGCTCAAGCTTACTTAAGAAAAAGTATCTTAC 1739
 Db 18750 GCTTTTGTGATGAAGAAAAAGTTGCAAGCTCAAGCTTACTTAAGAAAAAGTATCTTAC 18809
 Qy 1740 CTCATCTCCAGACCGCAATGTTAAAGCAATCCAACTGAGATGTCAGAGATATTT 1799
 Db 18810 CTCATCTCCAGACCGCAATGTTAAAGCAATCCAACTGAGATGTCAGAGATATTT 18869
 Qy 1800 ACAATCGTGAAGAGGAAAAAGCAATTTCCACTGCTGCACTTCATTAATGTTGAGC 1859
 Db 18870 ACAATCGTGAAGAGGAAAAAGCAATTTCCACTGCTGCACTTCATTAATGTTGAGC 18929
 Qy 1860 ATACAGTTGAGGTTAAAAACGTTAATTTGATTTCTCATTAAGATCATTAACATAATA 1919
 Db 18930 ATACAGTTGAGGTTAAAAACGTTAATTTGATTTCTCATTAAGATCATTAACATAATA 18989
 Qy 1920 TTAATTTGCTGTTGTTGATGATCAACATTAACAAAGCTCCAAATGCTTACCTTTGAG 1979
 Db 18990 TTAATTTGCTGTTGTTGATGATCAACATTAACAAAGCTCCAAATGCTTACCTTTGAG 19049
 Qy 1980 ATTTGTTTGCAGATTAAGTACTAGCTGAACACCTGACGAAGCTCCACTTCAATG 2039
 Db 19050 ATTTGTTTGCAGATTAAGTACTAGCTGAACACCTGACGAAGCTCCACTTCAATG 19109
 Qy 2040 ATGATGAGGAGCAATCCAGTGAAGATGTTTGAAGCAAGAAAGCAAGTGAAGTCCAA 2099
 Db 19110 ATGATGAGGAGCAATCCAGTGAAGATGTTTGAAGCAAGAAAGCAAGTGAAGTCCAA 19169
 Qy 2100 ATTAAGAACTTCAAGCGGATGAAGGCGAGTGAAGAAACACCTGCTGAGCAAGATTC 2159
 Db 19170 ATTAAGAACTTCAAGCGGATGAAGGCGAGTGAAGAAACACCTGCTGAGCAAGATTC 19229
 Qy 2160 CTCAGTGAAGCTGAAGAAAGTGAAGGCGCACTCAAGAAAGCAAGATTTTCTTGGCA 2219
 Db 19230 CTCAGTGAAGCTGAAGAAAGTGAAGGCGCACTCAAGAAAGCAAGATTTTCTTGGCA 19289
 Qy 2220 AAGTAAAGGATTTCTAGTCTGAAGGCGCAATGCAAGAACTCTAGCTGTTTACGAATA 2279
 Db 19290 AAGTAAAGGATTTCTAGTCTGAAGGCGCAATGCAAGAACTCTAGCTGTTTACGAATA 19349

OY	2280	ATTGACCTTCGAATTTTGGATAACAADAGATCATGTGCAGAAGCAAAAAAATTACTG	2339
DB	19350	ATTTACTCTTCACAAATTATGGATACCAATAGTATATGCGACGACAAAATAATTACTTG	1940
OY	2340	CGTTGTAAAAAGAGTAATCCTTCATCTGTAAAGTAGAAAAATAAAC	2369
DB	19410	CCTTGTAAAGAGTAATCCTTCATCTGTAAAGTAGAAAAATAAAC	19459
RESULT 11			
AF340221		2535 bp	DNA linear BCT 22-MAY-2001
LOCUS	AF340221	Streptococcus pneumoniae Phpa (phpa) gene, complete cds.	
DEFINITION	AF340221	Streptococcus pneumoniae Phpa (phpa) gene, complete cds.	
ACCESSION	AF340221		
VERSION	AF340221.1	GI:13447093	
KEYWORDS			
SOURCE		Streptococcus pneumoniae	
ORGANISM		Streptococcus pneumoniae	
		Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;	
		Streptococcus.	
REFERENCE		1 (bases 1 to 2535)	
AUTHORS		Zhang, J., Mast, A.W., Barniak, V., Mountzouras, K., Hostetter, M.K. and Green, B.A.	
TITLE		Recombinant Phpa protein, a unique histidine motif-containing protein from Streptococcus pneumoniae, protects mice against intranasal pneumococcal challenge	
JOURNAL		Infect. Immun. 69 (6), 3827-3836 (2001)	
MEDLINE		21246685	
LOCUS		11349048	
PUBMED		2 (bases 1 to 2535)	
FEATURES		Zhang, J., Mast, A., Barniak, V., Mountzouras, K., Hostetter, M. and Green, B.	
AUTHORS		Direct Submission	
TITLE		Submitted (25-JUN-2001) Department of Bacteriology, Wyeth Lederle Vaccines, 211 Bailey Road, West Henrietta, NY 14586, USA	
JOURNAL		Location/Qualifiers	
FEATURES		1..2535	
source		/organism="Streptococcus pneumoniae"	
		/mol_type="genomic DNA"	
		/db_xref="taxon:1313"	
gene		1..2535	
		/gene="Phpa"	
CDS		1..2535	
		/gene="phpa"	
		/note="histidine motif-containing protein"	
		/codon_start=1	
		/transl_table=1	
		/product="Phpa"	
		/protein_id="AAK6629.1"	
		/db_xref="GI:13447094"	
		/translation="IMKINKKYLVGSAAALILVSCSYELGLVQATYKENNRVSIVLSE OATOKTENLTPEDESVSKREGIAEOTIVKITDQGYVTSHGHYHYNNKVYDAIISKE LMMKPNTQLKDEDITISEIKGGYIVKDGKYTVLKQAHAADVNRTEENRQRQERS QHRSRGTPRNDDGAVALASOGRYTDDGYIFNADIIEDGDAYIPHGHHYIIPKN ELRSBELAAAKAPLSGRGNLSNSTRYSRONSDSRNWPSVPSTGTYTNSSNN TNSQASGNDNISLTKOLYKLPSRHVSPDLFDPDOTSRYPANGVAAPHGHYHF IYPSOLSLEEKRLARITPIRRYSNNHWVDSPDEQSPQSPPEPSPQAPNPQPAPS NPIDEKLYKENVKRVGCGGVYEENGVRVYIPAKLSMTAAGIDSKLAKDESJSHKLG AKKTDLPPSSDERFNKAIDLARIHODLDNKGQVDPEAUDNLERLKOVSSTKVL VDDTLAFLAPIRHPERLGPNAQTYYTDELOVKLAGKYTEDEGYIFDRDITSDEG DAYVTPMHTSHWTKDLSIEAEERAAAQAVYKEGGLPPSIDHSDSGNTAEAKAELAY NRYVAAKKVPLEDRMPYNLOYVEVNGSLIIIPHDYHLKEWFEGDKLEYAPGYSYL EDLATATYVYVENHNBRHSNGFGNMSDHQRKNQADPNQTERKNEEKPOETRE BETRREKRPQESRKPTPEEPDESBESSBPQVETEVKKLRLEADLLGK IQNVILSNMAKETILGLKNNLLPFEGTDNNITMAERKULALAESK"	
BASE COUNT	880 a	518 c	533 g 604 t
ORIGIN			

Query Match	60.0%	Score	1434.4	DB	1	Length	2535
Best Local Similarity	79.0%	Pred. No.	5.1e-296				
Matches 1724; Conservative	0	Mismatches	447	Indels	12	Gaps	1

QY	TTCTTACAGATTGGGACCTGTAATCAAGCTAGCAAGGTTAAGGAAAAATATCGTGTTCCTA	60
Db	TTCTTACAGATTGGGACCTGTAATCAAGCTAGCAAGGTTAAGGAAAAATATCGTGTTCCTA	119
QY	TATAGATTGAAAAACAAGCAGCAAAAAACGAGATTGACTCTGATGAGTTAGCA	120
Db	TATAGATTGAAAAACAAGCAGCAAAAAACGAGATTGACTCTGATGAGTTAGCA	179
QY	GGGTGAAGGAATCAATGCTGACCAATTCGTATCATCAGATAACAGACCAAGCTATGTCA	180
Db	GGGTGAAGGAATCAATGCTGACCAAAATTCGTATCAATAAACAAGCCAGCTATGTCA	239
QY	TTCAATGAGGACCACTATCATTTTATTCATAGTAAAGGTTCCCTATGACGCTATCATG	240
Db	TTCAATGAGGACCACTATCATTTATTCATAGTAAAGGTTCCCTATGACGCTATCATG	299
QY	TGAAGAAATTACTGAAAGATCCAAACTATAGCTAAAGATGAGATATTTGTTAATGA	300
Db	TGAAGAAATTGCTGATGAAGATCCAACTACACCTTAAAGGACGAGATATTTACAGTGA	359
QY	GGTCAAGGGTGGATATGTTATCAAGGTAGATGAAAAATCTATGTTTACCTTAAAGATGC	360
Db	AATCAAGGGTGGTATGATTTAAGTAAAGTAAATACATATGTTTACCTTAAAGATGC	419
QY	TGCCCCAGCGGATTAACGTCCGTACAAAAAGGAAATCAATGCACAAAAACAAGAGCATAG	420
Db	AGCTCATCGGATTAATGTCGTACAAAAAGGAAATCAATGCACAAAAACAAGAGCATAG	479
QY	TCAACATCGTAAAGTGGAACTCCAAAGAAACGATGAGTGTCTTGCTCGACAGCTTCGCA	480
Db	TCAACATCGTAAAGTGGAACTCCAAAGAAACGATGAGTGTCTTGCTCGACAGCTTCACA	539
QY	AGACGCTATACTACAGATGATGTTATCTTTAATGCTTTGATATCATAGAGATAC	540
Db	GGGACGCTACACCAAGATGATGTTATCTTTAATGCTTGTATCATTTGAAGATAC	599
QY	TGGATGATCTTATATGCTTCTCATTGAGATCATTAACATTAACCTTAAGATGAGTT	600
Db	TGGATGATCTTATATGCTTCTCATCGGAGCAATTAACATTAACCTTAAGATGAGTT	659
QY	ATCAGCTAGCAGATTGGCTGCTGAGAGAACCTTCTCTATCTGAGGAGAAATCTGTCAA	660
Db	ATCAGCTAGCAGATTAGCTGCTGCAAAAGCTTCTTATCTGAGGAGAAATCTGTCAA	719
QY	TTCAAGAACCTATCGCCGACAAAAATAGCGATTAACACTTCAAGAACAACTGGGTACCTTC	720
Db	TTCAAGAACCTATCGCCGACAAAAATAGCGATTAACACTTCAAGAACAACTGGGTACCTTC	779
QY	TGTAGCAATTCAGAGAACTACAAATCTAAACACAAGCAACAACGCAACTAACAGTCA	780
Db	TGTAGCAATTCAGAGAACTACAAATCTAAACACAAGCAACAACGCAACTAACAGTCA	839
QY	AGCAAGTCAAGTAAATGATGATGATGATGCTTTGAAAACAGCTCTACAAACCTGCTTGAG	840
Db	AGCAAGTCAAGTAAATGATGATGATGATGCTTTGAAAACAGCTCTACAAACCTGCTTGAG	899
QY	TCAAAGCATGTAGAAATCTGATGAGCTTGTCTTTGATCAGCACAAATCAACAGTCAAC	900
Db	TCAAAGCATGTAGAAATCTGATGAGCTTGTCTTTGATCAGCACAAATCAACAGTCAAC	959
QY	AGCTAAGAGTGTCCAGTCCACAGCAGAGATCATTAACACTTCAATCTCCTTAATCTCAAT	960
Db	CGCCAAAGTGTGTCTGTACCGCAGCAGACCATTAATCACTTTATCTTCAACT	1019
QY	GTCGAAATGGAAGAAAGGAATCGTCTGTAATATCCCTCGTATGTTCAACCAATG	1020
Db	GTCACCTTTGGAAAAAAATTTGCTGTATATTTCCCTTCGTATGTTCAACCAATG	1079
QY	GGTACACGATTTCAAGGCGAGAACCAACCAAGTCCACAACGCACTCCGAACTTACGAG	1080
Db	GGTACACGATTTCAAGACCAAGAACCAACCAAGTCCACAATGCACTCCGAACTTACGAG	1139
QY	CCGCAACCTGACCAAAATCTTAAATAGACTCAA-----ATTCTTCTTTGGT	1128

```

Db      1140 TCCGCAACCTGCACCAATCTCAACGACTCCAAACCAATTCATTTGATGAGAAATTTGCT 1199
      1129 TAGTACGCTGTAGCAAGAAAGTTGGGGAAGATATGATTCGAAGAAAGGCATCTCTCG 1188
      1200 CAAAGAGCTGTTCCGAAAGTAGGCGATGTTATGCTTTGAGGAAAGAGGAGTTCTCG 1259
      1189 TTATGCTTTGGGAAAGATTTTACATCTGAAACCTGTTAAATAATCTTGAAGCAAGTTATC 1248
      1260 TTATATCCAGCGAAGGATCTTTCAACAGAAACAGCAGCAGGATTTGATAGCAAACTGGC 1319
      1249 AAAACAGAGAGTGTTCACACCTTTAACTGCTAAATAAGAAATGTGCTCTCTGTA 1308
      1320 CAGAGAGGAAAGTTTATCTCATAGCTAGAGCTAAGAAACTGACCTCCCATCTAGTGA 1379
      1309 CCAAGATTTTATGATTAAGCAATATATCTGTTAACTGAGGCTCAATTAAGCCCTGTTGN 1368
      1380 TAGAGATTTTACAAATAGGCTTATGACTTACTAGCAAGAAATTCATCAAGATTTACTTGA 1439
      1369 AATTAAGGCTGTAATTTCTGATTTCCAGCCTTAGACAATTAATTAAGACGCTTGAATGA 1428
      1440 TATATAAGTGCACAAAGTATTTTGAAGCTTTGATTAACCTGTTGAACGACTCAAGA 1499
      1423 TGAATCGACTAATTAAGAAATAATGTTAGATGATTTATTTGGCATTTCTAGCAACCAATTAC 1488
      1500 TGTCTCAAGTGAATTAAGTCAAGTTAGTATGATTTCTTGCCTTCTAGCTCCGATTCG 1559
      1489 CCATCCAGAGCAGCTTTGGCAACCAATTTCTCAAAATTTGATGATCTGAAAGCAAGTTGC 1548
      1560 TCATCCAGAGCTTTAGGAAACCAATCGCAAAATTTACCTCACTGATGATGAGATTTGA 1619
      1549 TATTCGCTAATTTAGCTGATGATATATCAACGTCAGATGTTACATTTTGTGATGAACATGA 1608
      1620 AGTACCGAAGTTGGCAGGCAAGTACACACAGAAACGGTTATATTTTGTATCTCTGTA 1679
      1609 TATTAATCACTGATGAAGAGATGCATATGTAAACGCTCATATGGCCATATGATCATCTGAT 1668
      1680 TATTAACAGATGATGAGGGGATGCTATGTAATCTCAACATATGACCATATGACATGATGAT 1739
      1669 TGGAAAGATTAACCTTTCTGATTAAGAAATAAGTGAAGTCAAGCTTATCTTAAGAAAA 1728
      1740 TAAAAAGATATTTGTCTGAAGCTGAGAGCGGACCGGCTTATGCTTAAGAAAGAA 1799
      1729 AGTATCTTACCTCCATCTCCAGACGCAATGTTTAAAGCAATCCAACTGGAGATGATGTC 1788
      1800 AGGTTGACCTCTCTTTCAGACAGACATGATGAGAAATTAAGGAGCAAAAGAGAC 1859
      1789 AGCAGCTATTTTAACTGTCGTGAAGGGGAAAAAGCAATTTCCATCTGTCGATTCATATA 1848
      1860 AGAGCTATCTTCAACCGGCTGAAGCAGCTAAGAAAGTGCCATGATCGTATGCTTA 1919
      1849 TATGTTGAGCATACAGTTGAGGTTAAACGGTAAATTTGATTAATCTCTCATAGAGTGA 1908
      1920 CAATCTTCAATATACGTGAAGTCAAAAACGGTATTTAATCATCTCATATATGACCA 1979
      1909 TTACCAATATATTAATTTGCTTGTGATGATGATCAATCAAAAGCTCCAAATGGCTA 1968
      1980 TTACCAATATATCAATTTGAGTGTGTTGACGAGGCTTTTATGAGGCACTTAAGGGGTA 2039
      1969 TACCTTGAAGATTTGTTTGCAGCATTAAGTACTAGTGAACACCTTGAAGAAAGTGC 2028
      2040 TAGCTGTGAGATCTTTTGGCAGCTGTCAAGTACTATGTCGAACATCCAAAGAAAGTGC 2099
      2029 ACATTTTAATGATGATGGGCAATGCCAGTGAAGTGTGTTAGGCAAGAAAGCAACAG 2088
      2100 GCAATTCAGATTAATGTTTGTGTAAGCTTGAAGCCTATGTTCAAGAAACAAATAATGCTCA 2159
      2089 TGAAGTCCAAATTAAGAACTTCAAAAGCGATGAAGGCAAGTGAAGGAACAACCTGCTGA 2148
      2160 AGCTGATACATCAACGAGAAACCAACGAGGAAACCTCAGACGAAATAACCTGA 2219
      2149 GCCAGAGTCCCTCAAGTAGAGA 2171

```

```

Db      2220 GGAAAGAACCCCTCGAGAAAGA 2242

RESULT 12
AX343072      2523 bp      DNA      linear      PAT 12-JUN-2002
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
    source
        1..2523
        /organism="unidentified"
        /mol_type="genomic DNA"
        /db_xref="taxon:32644"
        /note="BVH-11"
BASE COUNT      879 a      523 c      526 g      595 t
ORIGIN
Query Match      57.5%; Score 1374.2; DB 6; Length 2523;
Best Local Similarity 73.9%; Pred. No. 3.7e-283;
Matches 1819; Conservative 0; Mismatches 539; Indels 102; Gaps 2;

1  TTCTTACGAGTTGGGACTGTATCAAGCTAAGAGCGTTAAGAAATAATCGGTTCTCTTA 60
60  TGTCTATGATCAAGATTTTGGCATCAAGCTCAAACTGTAAGAAATAATCGGTTCTCTTA 119
61  TATAGATGAAAAACAAGGACGCAAAAAACGAGAAATTTGACTCCTGATGAGTTAGCA 120
120  TATAGATGAAAAACAAGGACGCAAAAAACGAGAAATTTGACTCCTGATGAGTTAGCA 179
121  GCGTGAAGAAATCAATGCTGAGCAAAATGCTATCAAGATTAAGACCAAGGCTATGTCAC 180
180  GCGTGAAGAAATCAAGGCGCAAAACAAATGCTATCAAGATTAAGGATTAAGTGAAC 239
181  TTCAATGCGGACGCACTATCATTAATTAAGTGAAGTTCCTTATGAGCTATGATGAC 240
240  CTCTATGAGACATTAATCATTAATTAATTAAGTGAAGTTCCTTATGAGCTATGATGAC 299
241  TGAAGAAATTAATCAATGAAGATCCAAACTATTAAGCTTAAAGATGAAGATTAATGA 300
300  TGAAGAGCTCTCCATGAAGAAATCCGAATTAATGAAGTGAAGATTAATGAATGA 359
301  GGTCAAGGCTGATATGTTATCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 360
360  AATCAAGGCTGATATGTTATCAATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 419
361  TGCCCAAGGCTAATCGTCCGTAAGAAAGAAATTAATGAAGTGAAGTGAAGTGAAGTGA 420
420  AGCTATGCGGATTAATGTTCCGTAAGAAAGAAATTAATGAAGTGAAGTGAAGTGAAGTGA 479
421  TCAATGCTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 480
480  TCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 539
481  AGAGCGTATTAATCAAGATGATGTTATATCTTAATGCTTGAATGATGATGATGATGAT 540
540  GGGAGCTAACCAACAGATGATGTTATATCTTAATGATGATGATGATGATGATGATGAT 599
541  TGTGATGCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 600
600  GGGGATGCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 659
601  ATCACTGAGGAGTGTGCTGCTGCAAGGCTTCTATCTGCTGAGGAAATCTGTCAA 660

```

```

Db      660 ATCAGCTAGCGAGTGGCTGCTGCAGAAAGCCCTTCTATCTGTGTCGGGAAATCTGTCAA 719
Oy      661 TTCAAGAACTATGCGCCGCAAAATAGGATATACCTTCAGAGACAACTGGGATCTTC 720
Db      720 TTTAAGAACTATGCGCCGCAAAATAGGATATACCTTCAGAGACAACTGGGATCTTC 779
Oy      721 TGTAGCAATCCAGAACTACAAATATCTAACAGAGCAACACAGCAACACTACAGTCA 780
Db      780 TGTAGCAATCCAGAACTACAAATATCTAACAGAGCAACACAGCAACACTACAGTCA 839
Oy      781 AGCAAGTCAAAAGTATATGACATGATAGTCTCTTGAACAGCTCTACAACTGCTTTGAG 840
Db      840 AGCAAGTCAAAAGTATATGACATGATAGTCTCTTGAACAGCTCTACAACTGCTTTGAG 899
Oy      841 TCAAGGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
Db      900 TCAAGGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 959
Oy      901 AGCTAGAGGTGTGGCAGTGCACACAGGATCATACCACTTCACTCCCTTACTCTCAAT 960
Db      960 CGCCAGAGGTGTAGTGTGCTCCCTCATGTAACCATTCACACTTATCCCTTATGAACAAAT 1019
Oy      961 GTCTGATTTGGAGAACGATGCTGCTGATTAATCCCTGTTGTTATCGTTCAACCATTC 1020
Db      1020 GTCTGATTTGGAGAACGATGCTGCTGATTAATCCCTGTTGTTATCGTTCAACCATTC 1079
Oy      1021 GGATCAATTCAGAGCCAGAACCAAGTCCACACCGACTCCGGAAACCTAGTCCAGG 1080
Db      1080 GGATCAATTCAGAGCCAGAACCAAGTCCACACCGACTCCGGAAACCTAGTCCAGG 1139
Oy      1081 CCCGCAACCTGACCAAACTTTAAATAGACTCA-----ATCTCTCTTTGGT 1128
Db      1140 TCCGCAACCTGACCAAACTTTCAACAGCTCCAGCAATCCCAATGATGAGAAATGGT 1199
Oy      1129 TAGTACGCTGGTACGAAAGTGGGAGAGATATGATTCGAGAAAGGCGATCTTCG 1188
Db      1200 CAAGAAACCTGTTCCAAAGAGGCGATGATGATGATGATGATGATGATGATGATGATGAT 1259
Oy      1189 TTATGCTCTTGGAGAAATTTACATCTGAAACCTTTAAATCTTGAAGCAATTTATC 1248
Db      1260 TTATGCTCTTGGAGAAATTTACATCTGAAACCTTTAAATCTTGAAGCAATTTATC 1319
Oy      1249 AAAACAGAGAGTGTTCACACACTTTAACTGCTAAAAAGAAATGTTGCTCCTGTA 1308
Db      1320 CAAGAGAGAAAGTTATCTCAATAGCTAGAGCTAAGAAACTGACCTCCATCTAGTGA 1379
Oy      1309 CCAGAAATTTATGATTAAGCATATATCTGTTAACTAGAGCTCAAAAGCTTGTG 1368
Db      1380 TCGAGAAATTTATCAATTAAGGCTTATGACTTACTACAGAAATTCACCAAGATTTTACTGA 1439
Oy      1369 AAATTAAGGTGCTATTTCTGATTTCCAGCCTTACACAAATTTAGAACGCTTGATGA 1428
Db      1440 TAAATTAAGGTGCTATTTCTGATTTGAGCCTTGAATTAACCTGTTGAAACGCTCAAGA 1499
Oy      1429 TGAATCGACTATTAAGAAATTTGATGATGATTTATTTGCGATTCCTAGACCAATATC 1488
Db      1500 TGTCTCAAGTATTAAGTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1559
Oy      1489 CCATTCAGAGGCTTGGCAAACTTCAAAATGATGATGATGATGATGATGATGATGATGAT 1548
Db      1560 TCATTCAGAGGCTTGGCAAACTTCAAAATGATGATGATGATGATGATGATGATGATGATGAT 1619
Oy      1549 TATGCTCAATTTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1608
Db      1620 AGTAGCCAGGTGGAGGAGTACACACAGAAACGCTTATATCTTTGATGATGATGATGATGAT 1679
Oy      1609 TATTAATCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1668
Db      1680 TATTAATCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1739
Oy      1669 TGAAGAAATAGCCCTTTCTGATTAAGGAAAGGTTGACCTCAAGCTTACTTAAAGAAA 1728
Db      1740 TAAAGAAATAGTGTGCTGAAAGCTGAGAGAGCGGACCCAGGCTTATGCTTAAAGAGAA 1799

```

```

Oy      1729 AGGTATCCTTACCTCCATCTCCAGAGGAGATGTTAAAGCAATCCAACTGGAGATAGTGC 1788
Db      1800 AGGTTGACCCCTCCTTCGACAGACATCAGAGATTGAGAAATATCTAGGCAAAAGAGC 1859
Oy      1789 AGCAGCTATTTTACATGCTGTGAAAGGGGAAAAAGAAATTCACCTCGTTGCATTCATA 1848
Db      1860 AGAAGCTATCTACAAACCGCGGAAAGCAGCTTAAAGAGTGCCACTTGATCGTATGCCCTTA 1919
Oy      1849 TATGTTGAGATACAGTGGGTGTTAAACGGTAATTTGATTTTCTCTCATTAAGATCA 1908
Db      1920 CAATCTTCATATATCTGATGAGATCAAAAGGTAATTAATCAATCAATTAATGACCA 1979
Oy      1909 TTACCATTAATTAATTAATTTGCTGTTGATGATCAACATACAAAGCTCCAAATGCGTA 1968
Db      1980 TTACCATTAATTAATTAATTTGAGTGTGTTGAGAGGCTTTATAGGAGCTTAAGGGGTA 2039
Oy      1969 TACCTTGAGAGATTTGTTGCGAGATTAAGTACTAGTACAGACCTTGACGAACTGC 2028
Db      2040 TACTTGGAGGATCTTTTGGGCACTGCAAGTACTATGTCGAACTCAACGAAACGATCC 2099
Oy      2029 ACATTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2088
Db      2100 GCATTCAGATATATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2159
Oy      2089 TGAAGATCCAAATTAAGAACTTCAAGCGGATGAGAG----- 2125
Db      2160 AGCTGATACCAATCAAAACGAAAAACCAAGGAGAGAAACCTCAGACAGAAAACTGTA 2219
Oy      2126 ----- 2125
Db      2220 GGAAGAAACCCCTCGAGAGAGAAACCAAGGAGAGAAACAGAGTCTCCAAACCAAC 2279
Oy      2126 -----CCAGTAAGAGAAACACTGCTGAGCCAGAGTCCCTCAAGTACAGACTGAAA 2178
Db      2280 AGAGAAACCAAGAAAGATATCCAGAGATTCGAAAGACTGAGCTGAGCTGAGAA 2339
Oy      2179 AGTAAAGCCCACTCAAAAGAGAGAGATTTTCTGCGAAAGTAAAGGATTTAGTCT 2238
Db      2340 GGTGAAAGAAACCTGAGAGAGGCTGAGAGATTTCTTGGAAGAAATCCAGATCCAAATAT 2399
Oy      2239 GAAGCAATGCAACAGAACTAGCTGTTTACGAATTAATTTGACTCTTCAATATAT 2298
Db      2400 CAAGTCCAAATGCAAAAGAGACTTCACAGGATTAATAATTAATTTGACACCA 2459
Oy      2299 GGATTAACATATGATTCATGCGAGAGCAAGAAATTTCTGCGTTTAAAGGAGATTA 2358
Db      2460 GGACCAACAAATTAATTAATGCGAGAGGCTGAAAACTATTTGCTTTATTAAGGAGATTA 2519

RESULT 13
AX343073      2647 bp      DNA      linear      PAT 12-JAN-2002
LOCUS      AX343073
DEFINITION      Sequence 4 from Patent WO0198334.
ACCESSION      AX343073
VERSION      AX343073.1 GI:18152271
KEYWORDS
SOURCE      unidentified
ORGANISM      unidentified
unclassified.
REFERENCE
1      Hamel,J., Ouellet,C., Charland,N., Martin,D. and Brodeur,B.
AUTHORS      Streptococcus antigens
TITLE      Patent: WO 0198334-A 4 27-DEC-2001;
JOURNAL      SHIRE BIOCHEM INC. (CA)
FEATURES
source
1. .2647
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/note="BVH-11"
BASE COUNT      934 a      538 c      556 g      619 t
ORIGIN

```

Query Match 57.5%; Score 1374.2; DB 6; Length 2647;
 Best Local Similarity 73.9%; Pred. No. 3.7e-283;
 Matches 1819; Conservative 0; Mismatches 539; Indels 102; Gaps 2;

```

QY 1 TTCTTACGAGTTGGAGCTGATCAGCTTGAAACGGTTAAGAAAATATCGTGTCTTA 60
DB 104 TCGTTATGTAAGTCTGATGCTGATCAGCTGAACTGTGTAAGAAAATATCGTGTCTTA 163
QY 61 TATAGTGAAGAAACAAGCGCAAAAGGAAATTTGACTCCCTGATGAGTTAGCA 120
DB 164 TATAGTGAAGAAACAAGCGCAAAAGGAAATTTGACTCCCTGATGAGTTAGCA 223
QY 121 GCGTGAAGAAATCAATGCTGAGCAATCGTCATCAAGATPAAGACCAAGGCTATGTCAC 180
DB 224 GCGTGAAGAAATCAAGCGCAAAATCGTCATCAAGATTTAGCGATCAAGGTTATGTGAC 283
QY 181 TTCACATGGCGACCACTATCATTTATTAACAATGTAAGTTCCTTATGACGCTATCATCAG 240
DB 284 CTCTCATGAGACCAATTATCTATTAATGCAAGGTCCTTATGATGCATCATCAG 343
QY 241 TGAAGATTACTCATGAAGATCCAACTATAAGCTAAAGATGAGATATTGTAATA 300
DB 344 TGAAGATCTCTCATGAAGATCCGAATTATCAAGTTGAGGATTCAGACATTTGCAATGA 403
QY 301 GGTCAAGGCTGATATGTTATCAAGGATGAGAAATCTATGTTTACCTTAAGGATGC 360
DB 404 AATCAAGGCTGATATGTTATCAAGGATGAGAAATCTATGTTTACCTTAAGGATGC 463
QY 361 TGCCCAACGCGGATTAACGTCCTGTAACAAGAGAAATCAATGCACAAAAACAAGACATAG 420
DB 464 AGCTCATGGGATTAATGTCCTGTAACAAGAAATCAATCGGCAAAAACAAGACATAG 523
QY 421 TCAAGATCGTGAAGTGAACCTCCAAGAACGATGCTGTGCTTGGCTGGACGTTCCGA 480
DB 524 TCAGATCGTGAAGGAGGACCTTCAGCAAAACGATGCTGTGCTTGGACGTTCCGA 583
QY 481 AGGACGCTATCTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
DB 584 GGGAGGCTACACACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 643
QY 541 TGGTATGCTTAATATGCTTCTCTCATGAGATCATTAACATTAACATTCCTTAAGATGAGTT 600
DB 644 GGGCATGCTTATATGCTTCTCTCATGAGATCATTAACATTAACATTCCTTAAGATGAGTT 703
QY 601 ATCAGCTACGAGTTGGCTGCTGACAGAACCTTCTCATATGCTGCAAGAAATCTGTCAAA 660
DB 704 ATCAGCTACGAGTTGGCTGCTGACAGAACCTTCTCATATGCTGCGGAAATCTGTCAAA 763
QY 661 TTCAGAAACCTATCGCGCAAAATAGGATTAACACTTCAGAAACAACTGGGTACCTTC 720
DB 764 TTTAAGAACTTATCGCGCAAAATAGGATTAACACTTCAGAAACAACTGGGTACCTTC 823
QY 721 TGTAAAGCAATCCAGAACTACAATACTAACAACAAGCAACAAGCAACTAACAGTCA 780
DB 824 TGTAAAGCAATCCAGGAATCAACAATACTAACAACAAGCAACAAGCAACTAACAGTCA 883
QY 781 AGCAAGTCAAAGTATGATGATGATGATGCTTTGAAACGCTGTACAACTGCTTTGAG 840
DB 884 AGCAAGTCAAAGTATGATGATGATGATGCTTTGAAACGCTGTACAACTGCTTTGAG 943
QY 841 TCAACGACATGTAGATCTGATGGCTGTCTTTGATCCAGCAAAATACAAAGTCGAAC 900
DB 944 TCAACGACATGTAGATCTGATGGCTGTCTTTGATCCAGCAAAATACAAAGTCGAAC 1003
QY 901 AGCTAGAGGTGTGACAGTGCACACAGAGATCATTAACACTTATCCCTTACTCTCAAT 960
DB 1004 GCGCAAGGTGTAGCTGCTCCCTCATGTAACCATTAACACTTATCCCTTATGAACAAT 1063
QY 961 GTCTGAATGTGAAGAACGATGCTGTATTAATCCCTTCTGTTATGTTCAACCATTTG 1020
DB 1064 GTCTGAATGTGAAGAACGATGCTGTATTAATCCCTTCTGTTATGTTCAACCATTTG 1123

```

```

QY 1021 GGTACCAATTCACAGCCAGAACCAACAGTCCACACACGACTCCGGAACCTAGTCCAG 1080
DB 1124 GGTACCAATTCACAGCCAGAACCAACAGTCCACACACGACTCCGGAACCTAGTCCAG 1183
QY 1081 CCCGCACTTCGACCAAAATCTTAAATATGACTCAA-----ATTCTTTTGGT 1128
DB 1184 TCCGCACTTCGACCAAAATCTTCAACGAGCTCCAAAGCAATCCAAATGTATGAGAAATGGT 1243
QY 1129 TAGTACGCTGGTGAAGAAAGTTGGGGAAGGATATGTATTCGAAGAAAGGGGACTCTCG 1188
DB 1244 CAAGAACCTGTTCCAAAGATGAGGATGATGATGATCTTTGAGGAATGAGATTCTCG 1303
QY 1189 TTATGCTTTGGGAAGATTTACCATCTGAACCTGTTAAATCTTGAAGCAATGATC 1248
DB 1304 TTATATCCAGCAAGAAATCTTTGACGAGAAACACAGACAGGCAATGATGAGAAATCGGC 1363
QY 1249 AAAACAAAGAGTGTTCACACACTTAACTGCTTAAAAAGAAAAATGTTGCTCTGTGA 1308
DB 1364 CAAGCAAGAAAGTTATCTATAGCTAGGAGCTAAGAAAACTGACCTCCCATCTAGTGA 1423
QY 1309 CCAGAAATTTATGATAAGCATATATCTGTTAATGAGGCTCAATAAGGCTGTTGN 1368
DB 1424 TCAGAAATTTTCAATTAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACTGA 1483
QY 1369 AAATTAAGGCTGTAATCTGATTTTCCAGGCTTGAACAAATTAATAGAACGCTTGAATGA 1428
DB 1484 TAATTAAGTGAACAGTGAATTTTGAAGCTTGTGATTAACCTGTGGAACGACTCAAGA 1543
QY 1429 TGAATGCACTAATTAAGAAAAATGATGATGATTTATGCGATTCCTAGCACAATTAC 1488
DB 1544 TGTCTCAAGTGAATTAAGTCAAGTTGATGATGATTTCTTGCTTCTTACGCTCCGATTCG 1603
QY 1489 CCATCAGAGGACCTGGGCAAAACCAATTTCAATGATGATTAAGTGAAGAGAGTTCG 1548
DB 1604 TCATCAGAGGCTTGAAGAAAAACCAATGCGAAATTAACCTACATGATGATGAGAAATTC 1663
QY 1549 TATTCCTAATTAAGCTGATTAAGTATACAAAGCTCAGATGTTACATTTTGTATGAACATGA 1608
DB 1664 AGTACCAAGTTGGAGGCAAGTACACAGAACAGAGACGTTATATCTTTGATCCCTGCTGA 1723
QY 1609 TATTAATCAATGTAAGAGAGATGATGATGTAACGCTCATATGAGGCAATGACTGAT 1668
DB 1724 TATAACCAATGTAAGGAGGATGCTTATGTAACCTCAATATGACCCATAGCACCTGAT 1783
QY 1669 TGGAAATATAGCCTTTCTGATTAAGGAAAAAGTTGCACTCAAGCTTATCTAAAGAAA 1728
DB 1784 TAAAAAATATGTTGTGTGAAGCTGAGAGCGGACCCAGGCTTATGCTTAAAGAGAA 1843
QY 1729 AGGTATCTACCTCCATCTCCAGAGCAGATGTTAAAGCAAAATCCAACTGGAGATAGTC 1788
DB 1844 AGGTTGACCTCTCTTTCAGACAGACCATCAGGATTCAGGAAATCTGAGGCAAAAGAGC 1903
QY 1789 AGCAGCTATTTAACAATCGTGTGAAGGGGAAAAAGAAATTCACACTGTTGCACTTCATA 1848
DB 1904 AGAAGCTATTTAACAACCGGTGAAGAGCTTAAGAAAGGTCACACTGATGATGCTTA 1963
QY 1849 TATGTTGAGCATACAGTGAAGTTAAAAAGGTTAATTTGATTTCTCATAGAGATCA 1908
DB 1964 CAATCTTCAATATATCTGTAAGATCAAAAAAGGTTAATCAATCACTCATATGAGCA 2023
QY 1909 TTACCATTAATTAATTTGCTGTTGATGATCACAATCAAGGCTCCAAATGGGCTA 1968
DB 2024 TTACCATTAATTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2083
QY 1969 TACCTTGAAGATTTGTTGGACAGATTAAGTACTACAGTGAACACCTGACGAGCTGC 2028
DB 2084 TACTCTTGAAGATTTGTTGGACAGTGAAGTACTACAGTGAACACCTGACGAGCTGC 2143
QY 2029 ACATTCATATGATGATGAGGAGCATGCGAGTGAAGATGATGATGATGATGATGATGATGAT 2088
DB 2144 GCATTCAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2203
QY 2089 TGAAGATCCAAATGAAGACTTCAAGCGGATGAAGAG----- 2125

```



```

Db      1389  TCGACAAGTTGATTTTGGAGCTTTGGATAAAGCTGTTGAAACGATCAGATGCTCAAG 1448
Qy      1438  TAATAAGAAAAATTTGGTAGATGATTTATTTGGCATTCCTAGCAACCAATTACCATCCAGA 1497
Db      1449  TGATAAAGCAAGTTAGTGAAGATTTCTTGCTTTCTTAGTCCGATTCGTCATCCAGA 1508
Qy      1498  GCGACTTGGCAAAACCAATTCTCAATTGAGTATGTAAGAGCAAGATTGTAATGCTCA 1557
Db      1509  ACGTTAGAAAAACCAATGCGCAATTAACCTACAGTATGATGATGATCAAGTACCA 1568
Qy      1558  ATTACTGATAGTATACAAAGCTCAGATGTTATCAATTTTGTGATGAACATGATTAATCG 1617
Db      1569  GTTGGAGGCAAGTACAAACGAAACGTTATATCTTTGATCTCTCGAATATAACCG 1628
Qy      1618  TGATGAAGAGATGATGATTAAGCGCTATATGAGGCGCATAGCTACTGATGTAAGAAAAA 1677
Db      1629  TGATGAGGGGATGCTTATGTTACTCAATATGACCCATAGCCATGGAATTAAGAAAAA 1688
Qy      1678  TAGCTTTCTGATTAAGAAAAAGTTGCAGCTCAAGCCTATACTAAAGAAAAAGTATCCT 1737
Db      1689  TAGTTTGTCTGAAGCTGAGAGAGCGGCGACCCAGGCTTATGCTAAAGAAAGTTTGAC 1748
Qy      1738  ACCCTCATCTCCAGACGAGATGTTAAAGCAATCCAACTGGAGATGTCAGACGCTAT 1797
Db      1749  CCTCTCTTGACAGACATCAGGATTCAGAAATATCGAGCAAAAGAGCAGAAAGCTAT 1808
Qy      1798  TTACATCGTGTGAAGGAAAAAGCAATTCCTGCTGCTGCACTTCATATATGGTTGA 1857
Db      1809  CTACACCGCGTGAAAGAGCTTAAGAGGTGCACCTTATGCTTATGCTTACATTTTCA 1868
Qy      1858  GCATACAGTTGAGTTAAAAACGTTAATTGATTTATCTCTCATTAAGATCATTAACATA 1917
Db      1869  AATATAGTAGAAGTCAAAAACGGTAGTTAATCATACTCATTAATGACCATTAACATA 1928
Qy      1918  TATTAATTTGCTGTTGTTGATGATCACAATCAAAAGCTCCAATGCTATACCTTGA 1977
Db      1929  CATCAAAATTTGAGTGGTTGAGCAAGGCTTTATGAGGACCTTAAGGGGTATATCTTTGA 1988
Qy      1978  AGATTGTTGGAGCATTAAGTACTAGTAGAACCCCTGAGCAACGTCACATTCTTA 2037
Db      1989  GGATCTTTTGGCGACTGTCAAGTACTATGTCAAACTCAAAACGAAACGTCGATTCAGA 2048
Qy      2038  TGATGATGGGCAATGCCAGTGAAGCTGTTAGGCAAGAAAGACCAAGTGAAGATCC 2097
Db      2049  TATGCTTTTGGTAAACGTTAGGACCATGTTCAAGAAACAAATAATGTCAGCTGATAC 2108
Qy      2098  AATTAAGAACTTCAAAACCGGATGA----- 2121
Db      2109  CAATCAAAACGAAAAACCAAGGAGAGAAACCTCAGACAGAAAAAAGCTGAGAAAGAAC 2168
Qy      2122  -----AGAGCCAGT 2130
Db      2169  CCCTGAGAGAGAAACCGCAAAACGAGAAACCAAGTCTCCAAAAACCAACAGAGAAAC 2228
Qy      2131  AGAGGAAACACCTGCTGAGCAGAAAGTCCCTCAAGTAGAGACTGAAAAAGTAAAGCCCA 2190
Db      2229  AGAAGATACAGAGAGATCAGAAAGACCTCAGCTCGAGACTGAAAAAGCTTGAAGAAA 2288
Qy      2191  ACTCAAGAGAGAAAGTTTGTCTGCGAAAGTAAAGCAATTCAGTTCGAAAGCCAAATGC 2250
Db      2289  ACTGAGAGAGCTGAAGATTTTACTTGGAAAAATCCAGATTCCAATTAATCAATTCATATC 2348
Qy      2251  AACCAAACTAGCTGCTGTTACGAATAATTGACTCTTCAATTATGATTAACAAATG 2310
Db      2349  CAAAGAGACTTCACAGAGATTAATAATAATTAATTAATTTGACAGCCAGAGAACCAATAC 2408
Qy      2311  TATCATGCGAGAGAGAAAAATTAATCTGCTGTTTAAAGAGATTA 2358
Db      2409  TATTAATGCGAGAGAGTGAAGAACTATGCTTTATTAAGAGAGATTA 2456

```

RESULT 15
AF318954

```

LOCUS      AF318954                2457 bp    DNA        linear    BCT 11-FEB-2001
DEFINITION Streptococcus pneumoniae pneumococcal histidine triad protein B
ACCESSION  AF318954
VERSION    AF318954.1 GI:12744741
KEYWORDS   Streptococcus pneumoniae
SOURCE      Streptococcus pneumoniae
ORGANISM   Streptococcus pneumoniae
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1 (bases 1 to 2457)
            Adamou,J.E., Heinrichs,J.H., Erwin,A.L., Walsh,W., Dormitzer,M.,
            Dormitzer,M., Dagan,R., Brewah,Y.A., Baxren,P., Lathigra,R.,
            Langemann,S., Koenig,S. and Johnson,S.
            Identification and characterization of a novel family of
            pneumococcal proteins that are protective against sepsis
            Infect. Immun. 69 (2), 949-958 (2001)
JOURNAL     21101045
MEDLINE     11159990
PUBMED      11159990
REFERENCE   2 (bases 1 to 2457)
            Adamou,J.E., Heinrichs,J.H., Erwin,A.L., Walsh,W., Dormitzer,M. and
            Johnson,S.
            Direct Submision
            Submitted (03-NOV-2000) Molecular Microbiology, MedImmune, Inc., 35
            West Walkins Mill Road, Gaithersburg, MD 20878, USA
FEATURES             location/Qualifiers
     source           1..2457
                     /organism="Streptococcus pneumoniae"
                     /mol_type="genomic DNA"
                     /serotype="4"
                     /db_xref="taxon:1113"
                     1..2457
                     /gene="phtB"
                     1..2457
                     /gene="phtB"
                     /note="phtB, protective pneumococcal antigen"
                     /codon_start=1
                     /transl_table=11
                     /product="pneumococcal histidine triad protein B
                     precursor"
                     /protein_id="AAK06759.1"
                     /db_xref="GI:12744742"
                     /translation="MKINKYLAGSVAIVALSVCSEYELGRYOAGODKESNRYAVTDG
                     DOAGKAEMLTPDEVSREKGINAEIOVLIKIDGCGVTSHGHHYHNVKVPYDAIISE
                     ELMDKPNYQKDSPIVNEIKGYIYKNGKYVYLKRAAHADNIRTKETIKRQOER
                     SHNHSRADNVAAARAQGRYTTDDGYTFNASDIIEDTGDAIVPHGHHYIIPNEL
                     SASLEAAEAAYNKGQGRSPSSSYNAPQPRLESHNHLVTPTVQNOENISL
                     LRELAKPLSERHVESDGLIFDPAQITSRARGAVVPHGNHYHFIPEQMSLEKRIA
                     RILPLRYNSHWVPDSRPEPSPOPTPEPSPOPAENPIDGKLVKEAVRVGQGV
                     FEENGVSRYITPAKDISAETAGIDSKLAKOSLSHKIGTKXTDLPSSDRFPYKAYDL
                     LARIHODLIDNKGRQVDFEALDNLERKDVSSDVKYVEDILAFATIRHPERLQKP
                     NAOITYTDEIDVAKLAGYTAEDGYITDPDITSDBEDAVYTPMTSHWTKDQSL
                     EAERAAQAVAXEKGLTPSTDHDSGTEAKGAIAVNRVKAAPKVLDRPYNLYQ
                     TVEVANGSLIIPHYDHNIKFEWDEBLYEAPKYTLDELATAYKYVLEHNEKPS
                     DNGPCNASDHYORANKGOADNONGTPEEPKPEEKPEETPEEPKPEEKPEEKPEKPT
                     BEPSEPESEPEPEVETEKVEKLEAREADLLGKIDPILIKSNAKETLGLKNNLLFGT
                     QDNNTYMAEKLLALKESK"
     sig_peptide      1..87
                     /gene="phtB"
                     247..264
                     /gene="phtB"
                     /note="Region: histidine triad"
                     616..633
                     /gene="phtB"
                     /note="Region: histidine triad"
                     718..792
                     /note="Region: histidine triad"
                     940..957
                     /note="Region: coiled-coil domain"
     misc_feature     /gene="phtB"
                     /note="Region: histidine triad"
                     1396..1488
                     /gene="phtB"
     misc_feature     /gene="phtB"
                     /note="Region: histidine triad"

```

```
misc_feature      /note="Region: coiled-coil domain"
                  /gene=".1674
                  /note="phtb"
misc_feature      /note="Region: histidine triad"
                  /gene=".1926
                  /note="phtb"
repeat_region     /note="Region: histidine triad"
                  /rpt_type=direct
                  /rpt_type=direct
repeat_region     /rpt_type=direct
                  /rpt_type=direct
BASE COUNT      832 a 496 c 533 g 593 t 3 others
ORIGIN
Query Match      41.5% Score 991; DB 1; Length 2457;
Best Local Similarity 65.7%; Pred. No. 2.9e-201;
Matches 1609; Conservative 1; Mismatches 697; Indels 141; Gaps 6;
Oy 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAACGGTTAAGGAAA---TAAATCGTGTTC 57
Db 60 TTCCATAGAGCTTGACCTTACCAAGCTGTGACGATTAAGAAAGCTTAATCGAGTTGC 119
Oy 58 CTATATAGATGAAACAGCGCAACGAAATTTGACTCTGATGAGTTAG 117
Db 120 TTATATAGATGTGTATCAGGCTGTCAAAAGCAGAAAATTGACACAGATGAAGTCAG 179
Oy 118 CAAGCGTAGAGAAATCATGTGAGCAATGCTGATCAAGATTAACAGACCAAGGCTATGT 177
Db 180 TAAAGGAGAGGAGATCAACGCCGACCAAAATTTGTATCAAGATTAAGATCAAGTTATGT 239
Oy 178 CACTTCACATGGCGCACTATCATTTATTAAGATGTAAGTTCTTATGACGCTATCAT 237
Db 240 GACCTCTGATGAGACATTAATCATTAATTAAGCAAGTTCTTATGATGACATCAT 299
Oy 238 CAGTGAAGAATTACTCATGAAAGATCCAACTATTAAGCTAAAGATGAGATATTGTTAA 297
Db 300 CAGTGAAGAAGCTCCCATGAAAGATCCGAATTAATCAATGAAGATTCAGACATTTGTA 359
Oy 298 TGAGGTCAAGGCTGTATGTTATCAAGGTGATGGAATAATCTATGTTTACCTTAAGGA 357
Db 360 TGAATTAAGGCTGTATGTTATCAAGGTGATGGAATAATCTATGTTTACCTTAAGGA 419
Oy 358 TGCTGCCACGCGGATTAACGTCCTGTAACAAAGAGAAATCAATCGACAAAACAGAGCA 417
Db 420 TGCAGCTATGCGATTAATATTCGACAAAGAAAGATTAAGCTCGAAGACAGAGAG 479
Oy 418 TAGTCAATCATGTAAGGTGAATCTCAAGAAACGATGCTGTGCTTGGCAGCTTC 477
Db 480 CAGTCATATATCAT-----ACCTCAAGACAGATATATGCTGTGCTGACGACAGC 530
Oy 478 GCAAGGAGGCTATATAGATGATGTTATCTTTATGCTTGTGATATCATAGAGA 537
Db 531 CCAAGGAGGTTATACAGGATGATGAGATATCTTCAATGATCATGATATCATTAAGAGA 590
Oy 538 TACTGTGATGCTTATATGTTCTTCAATGAGATCATTAACATTAATCTTCAAGATGA 597
Db 591 CACGGGTATGTTATATGTTCTTCAATGAGATCATTAACATTAATCTTCAAGATGA 650
Oy 598 GTTATCAGCTAGCGAGTTGGCTGCTGCAAGAACCTTCTTATCTGTCAGAGAAATCTGTC 657
Db 651 GTTATCAGCTAGCGAGTTGGCTGCTGCAAGAACCTTATGGAATG----- 695
Oy 658 AAATTCAGAAACCTATGCCCAACAAATAGCATTAACCTTCAAGAAACAAATGGGTACC 717
Db 696 -GAAGCAGGAGATCTGCTCTTCTTCAAGTTCTAATTAATGCAATCAGCTCAACCA 754
Oy 718 TTCTTGAAGCATCAGAGAACTAAATTAACAAGCAACAAACAACTAAGAG 777
Db 755 GATTTCTAGAGAAACCAATCT-----GACTGTCACTCCAACTTA 794
Oy 778 TCAAGCAAGTCAAAATGATGATGATGATGCTTTGAAACAGCTCTCAAACTGCTTT 837
Db 795 TCATCAAAATCAAGGAGAAACATTTCAAGCTTTTACGTGAATGTATGCTAAACCTT 854
```

```
Oy 838 GAGTCAAGCATGTAGAAATCTGATGCGCTTGTCTTTGATCCAGACAAATCAGAAAGTC 897
Db 855 ATCAAGAGCATGTGAAATCTGATGCGCTTATTTGACCCAGCGCAATCAGAAAGTCG 914
Oy 898 AACAGCTAGAGTTGTGACGTCCACAGAGATCATTTACCATTTACCTTACTCTCA 957
Db 915 AACCGCAGAGTTACTGTGCTCCCTCAATGTAACATTAACCTTATCCCTTATGAACA 974
Oy 958 AATGCTGAATTTGAAGAGAAATGCTGATTAATTCCTTGTATGCTTGAACCA 1017
Db 975 AATGCTGAATTTGAAGAGAAATGCTGATTAATTCCTTGTATGCTTGAACCA 1034
Oy 1018 TTGGGTACCAAGATTCAGAGCGAGAAACCAAGTCCACACCGACTCCGAACTAGTCC 1077
Db 1035 TTGGGTACCAAGATTCAGAGCGAGAAACCAAGTCCACACCGACTCCGAACTAGTCC 1094
Oy 1078 AGGCCCCGCAACTGACCAAAATCTTAAATTAAGCTCAAAATTTCTTTGTTAGTACGT 1137
Db 1095 AAGTCCGCAACAGCTCCAGCAATCCAAATTAATGAGGAA-----ATTGCTCAAGAAAGC 1148
Oy 1138 GGTACGAAAGTTGGGAGAGATGATTAATCGAAGAAAGGAGCATCTCTGTATGCTT 1197
Db 1149 TGTTCGAAAGTACGAGATGATTAATGCTTTGAGAGAGATGAGTTCTCTTAATGCC 1208
Oy 1198 TGCAGAAAGATTTTACATCTGAAACTGTTAAAAATCTTGAAAGCAATTAACAAACAGA 1257
Db 1209 AGCCAGAGATCTTACAGAGAAACAGCAGCAGGATTAATGACAACTCGCCAGAGCAGGA 1268
Oy 1258 GAGTGTTCACACATTTAACTGCTAAAGAAAGATGTTGCTCTGTGACCAAGATT 1317
Db 1269 AAGTTTATCTCATAGCTAGAGAACTAAGAAAGTCACTCCCATCTGATGATTCAGAAAT 1328
Oy 1318 TTATGATTAAGCAATATATCTGTAATCAAGGCTCATTAAGCCTTGTGTAATTAAGG 1377
Db 1329 TTCAATTAAGGCTTATATCTTACTAGAGAAATTAACCAAGATTAATTAATTAAGG 1388
Oy 1378 TCGTAATTCGATTTTCAAGGCTTACCAAAATTTTGAAGCGTGAATGATGATGATGAC 1437
Db 1389 TCGCAAGTGTATTTGAGGCTTGAATACCTGTTGAAGACATCAAGAGATGCTCAAG 1448
Oy 1438 TAAATTAAGAAATTTGATGATGATTTATGAGCATCTTCAAGCAATTAACCAATCCAGA 1497
Db 1449 TGATTAAGTCAAGTATGAGAAATTTCTTCTTCTGATGCTCGATTCATCCAGA 1508
Oy 1498 GCGACTTGGCAACCAATTTCTCAATGAGTATCAAGAGAGAGAGTGTATGCTCA 1557
Db 1509 ACGTTTGAAGAAACCAATGCGCAAAATTAACCTACATGATGATGATTAAGTACCA 1568
Oy 1558 ATTAGCTGATTAAGTATCAACGTCAGATGTTTACATTTTGAATGATGATGATTAATTCAG 1617
Db 1569 GTTGGCAGCAGATGACAGAGAGAGGCTTATATCTTTGATCTCTGTGATTAACAG 1628
Oy 1618 TGATGAAGAGATCATATGTAACGCTCATATGAGCCATATGTCATGATGATGATGAAGA 1677
Db 1629 TGATGAGAGGAGATCTATGTAATCTCAATATGACCAATGACCACTGATTAAGAAAGA 1688
Oy 1678 TAGCTTTTCTGATTAAGAAAGTTGAGCTCAAGCTTATATCAAGAAAGATGATCTCT 1737
Db 1689 TAGTTTGTCTGAAGCTGAGAGAGGAGGAGCCAGGCTTATGCTTAAGAAAGATGATGAC 1748
Oy 1738 AACTTCATCTTCAGAGCAGATGTTAAAGCAATCCAATGAGATGATGATGATGATGATGAT 1797
Db 1749 CCTTCTTCAGACAGATCATGATTAAGAAATTAATGAGGAAAGAGAGAGAGATGAT 1808
Oy 1798 TTACAAATGTTGTAAGGAGGAGAAAGCAATTCACGCTTGTGACTTCCATTAATGATGTA 1857
Db 1809 CTACAAACGAGTGAAGCAGCTAAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1868
Oy 1858 GCATACAGTTGAGTTAAAGCGTAAATTTGATTAATCTTCAATGATGATGATGATGATGATGAT 1917
Db 1869 ATATACGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1928
```

QY 1918 TATTAAATTTGCTGTTGATGATCAACATCAAAAGCTCCAAATGCTATACCTTGA 1977
 Db 1939 CATCAATTTGAGTGTGTTGACGAGCCCTTATGAGGACCTAAGGGGTATATCTTGA 1988
 QY 1978 AGATTTGTTGGAGCATTAAGTACGTAGAACCCCTGACGAAAGCTCCACATTCTTA 2037
 Db 1988 GGATCTTTGGGAGCTGCTCAAGTACTATGTCGAACCTCAAGAGAGCTCCGATTCTGA 2048
 QY 2038 TGATGATGGGGCAATGCGAGTACGATGTGTTAGCAGAAAAGACCAAGTGAAGATCC 2097
 Db 2049 TATATGTTTGGTAACGTACGACCATGTTCAAAAACAAAATGTGCAAGCTGATAC 2108
 QY 2098 AAATAGAACTTCAAAAGCGATGA----- 2121
 Db 2109 CAATCAAAACGAAAAACCAAGCGAGAGAAACCTCAGACAGAAAAAAGTGAAGAAAC 2168
 QY 2122 -----AGAGCCAGT 2130
 Db 2169 CCCTCGAGAGAAACCGCAAAACGAGAACCAAGTCTCCAAAAACCAACAGAGAACCC 2228
 QY 2131 AGAGGAAACACCTGCTGAGCCAGAAAGTCCCTCAAGTACGAGTGAAGAAAGTGAAGCCCA 2190
 Db 2229 AGAAGATACCAAGAGGATCAAGAAACCTCAGCTCGAGACTGAAAGTTGAAAGAAA 2288
 QY 2191 ACTCAAGAGCAAGAAAGTTGCTGCGAAGTAAAGATTTAGTCTGAAAGCCAAATGC 2250
 Db 2289 ACTGAGAGAGGCTGAAGATTTACTTGGAATAATCCAGATTCATATCAAGTCCAAATGC 2348
 QY 2251 AACAGAACTAGCTGTTTACGAATAATTTGACTCTTCAATATATGATPAACATATG 2310
 Db 2349 CAAAGAGACTTCACAGATTAATAAATAATTAATTAATTTGACACCCAGAGCAACAAATAC 2408
 QY 2311 TATCATGCGAGAGCAAGAAATAATTAATCTGCTGTTTAAAGAGATTA 2358
 Db 2409 TATTATGCGAGAGCTGAATAAATAATTAATGCTTTATTAAGAGAGATTA 2456

RESULT 16
 AF318955 2517 bp DNA linear BCT 11-FEB-2001
 LOCUS AF318955
 DEFINITION Streptococcus pneumoniae pneumococcal histidine triad protein D precursor (phtd) gene, partial cds.
 ACCESSION AF318955
 VERSION AF318955.1 GI:12744743
 KEYWORDS
 SOURCE Streptococcus pneumoniae
 ORGANISM Streptococcus pneumoniae
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
 REFERENCE 1 (bases 1 to 2517)
 AUTHORS Adamou,J.E., Heinrichs,J.H., Erwin,A.L., Walsh,W., Gayle,T., Dormitzer,M., Dagan,R., Brewah,Y.A., Barren,P., Lathigra,R., Langermann,S., Koenig,S. and Johnson,S.
 TITLE Identification and characterization of a novel family of pneumococcal proteins that are protective against sepsis infect. Immun. 69 (2), 949-958 (2001)
 JOURNAL 2101045
 PUBMED 11159990
 MEDLINE 2 (bases 1 to 2517)
 REFERENCE Adamou,J.E., Heinrichs,J.H., Erwin,A.L., Walsh,W., Dormitzer,M. and Johnson,S.
 AUTHORS
 TITLE Direct Submission
 JOURNAL Submitted (03-NOV-2000) Molecular Microbiology, MedImmune, Inc., 35 West Watkins Mill Road, Gaithersburg, MD 20878, USA
 FEATURES
 source
 1..2517
 /organism="Streptococcus pneumoniae"
 /mol_type="genomic DNA"
 /seotype="4"
 /db_xref="taxon:1313"
 1..>2517
 /gene="phtd"
 1..>2517
 CDS

/gene="phtd"
 /note="phtd; protective pneumococcal antigen"
 /codon_start=1
 /transl_table=11
 /product="pneumococcal histidine triad protein D precursor"
 /protein_id="AAK06760.1"
 /db_xref="GI:12744744"
 /translation="MKINKYLAASVAVALVCSYELGRHAGOVKESRNVSYIDG
 DAGOKAENLTPDEVSKEGINABOIVIKIDGQVYVSHGHYVYNNKQVDAISL
 ELMKDPYQIKSDIYNEIKGVYIKDGRYLYLKAADINRTKEEIKRQOE
 SHNFGGSDQVAARAQGRYTTDDGTFVNASDIEDTGAIVLPDHGHVYLPKE
 LSASLAAEAVYMGKQSRSSSYNANAPRLSENNMLVPTVHONQGENISS
 LLRELVAKPLSERHVESDGLFDEPAQITSPARGVAPHGNNHFI PYEOMSELEKRI
 ALITPLRYSNNHWVDSRPEQPOSTPEPPOPAPNPQAPSNPIDEKLVKEAVR
 KVGQVFEENGVSRYTPAKDLSAETAGISLAKOBSLSHKLGAKKTDLPSSDREF
 YNKAYDLARIHODLNDKRGQVDFEALNDLBERKDVPSDKVLVDILAFIARH
 PERLGPNAQITTTDDETQVALKAGKYTTEDGYIFDPDITSDGDAVVTPLMTHSW
 IKDLSLEAEBAQAQAVAKERGLTPSPSDHDSGTEAKGAELIYRNVAAKRYLDR
 MPYNIQYVEVYNGSLIPYDHYNIKEFWFDEGLYEPKGYLTLEDLATVXYVEH
 PNERPHSDNGNMSDHYRKNKVDQDSKPEDKEHDEVEPTHPESDEKENAGLNP
 ADNLKYPSTDPHEEERAEEDTDEAEIPOVENSYVNAKIADAELAEKVTDPSTROMA
 METUTGLKSSILITGKDNNTISAEDSLALIKESQAPIC"
 1..87
 /gene="phtd"
 247..264
 /gene="phtd"
 /note="Region: histidine triad"
 418..485
 /gene="phtd"
 /note="Region: coiled-coil domain"
 619..636
 /gene="phtd"
 /note="Region: histidine triad"
 943..960
 /gene="phtd"
 /note="Region: histidine triad"
 1417..1509
 /gene="phtd"
 /note="Region: coiled-coil domain"
 1678..1695
 /gene="phtd"
 /note="Region: histidine triad"
 1930..1947
 /gene="phtd"
 /note="Region: histidine triad"
 526..717
 /rpt_type=direct
 1597..1779
 /rpt_type=direct
 BASE COUNT 830 a 509 c 545 g 633 t
 ORIGIN
 Query Match 41.4%; Score 990; DB 1; Length 2517;
 Best Local Similarity 68.1%; Pred. No. 4.7e-201;
 Matches 1478; Conservative 0; Mismatches 636; Indels 57; Gaps 5;
 QY 1 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAACGTTAAGAAAA---TAACTGCTTTC 57
 Db 60 TTCTTATTAAGTGTGCTGCTCCACCAAGCTGTCAGGTTAAGAAAGATCTAATCGAGTTTC 119
 QY 58 CTATATAGATGGAAGAAACAGAGCAAGCAAAACCGAGAAATTTGACTCCGATGAGGTTAG 117
 Db 120 TTATATATAGTGTATGCTAGCTGTCTCAAAAGGACAGAAACCTTGACACAGATGAAGTCTG 179
 QY 118 CAAGCGTAGAGATCAATGCTGAGCAATATGCTCAAGATTAACAGACCAAGGCTATGT 177
 Db 180 TAAAGAGGAGGAGATCAAGCGCAACATATGTCATCAAGATTAAGATCAAGTTATGT 239
 QY 178 CACTTCAATGGCGACCACTATATTAATTAAGTGAAGTTCTTATGACGCTATCAT 237
 Db 240 GACCTCTATGAGAGCAATTAATTAATTAAGTGAAGGCTCTTATGATGATCATAT 299

QY 238 CAGTGAAGAACTTACTCATGAAGATCCAACTATAGCTAAAGATGAGATATTGTAA 297
 Db 300 CAGTGAAGAGCTCCTCATGAAGATCCGAATTATAGTGAAGATTCAGACATTGTCA 359
 QY 298 TGAGTCAAGGGTGGATATGTTATCAAGTATGATGAAAATATCTATGTTTACTTAAGA 357
 Db 360 TGAATCAAGGGTGGTATGTTATCAAGTATGATGAAAATATCTATGTTTACTTAAGA 419
 QY 358 TGCTGCCACCGGTAACGTCCGTAACAAGAGGAATCAATGACAAAGAAAGACA 417
 Db 420 TGACGCTATCGGTAATATTCGACAAAGAAAGATTAACGTCGAAGACGAAACA 479
 QY 418 TAGTCACATCGTGAAGTGAACCTCAAGAACGATGCTGTGCTTGCCCTTGACGCTTC 477
 Db 480 CAGTCATATACCGGGGGTGTCT-----AACGATCAAGCAGTATGTCAGACGACAGC 533
 QY 478 GCAAGACGCTATATACAGATGATGTTATATCTTTATGCTTGTATCATAGAGA 537
 Db 534 CCAAGGACGCTATACAGATGATGTTATATCTTTATGCTTGTATCATAGAGA 593
 QY 538 TACGCTGATGCTTATGCTTCTCATGAGATCATTAACATTAATCTTCTTAAGATA 597
 Db 594 CACGGTATGCTTATATGCTTCTCATGAGACCATTAACATTAATCTTCTTAAGATA 653
 QY 598 GTTATCAGCTAGCGAGTGGCTGCTGACAGAACCTTCTTATCTGATCGAGAAATCTGTC 657
 Db 654 GTTATCAGCTAGCGAGTGGCTGCTGACAGAACCTTCTTATCTGATCGAGAAATCTGTC 688
 QY 658 AAATTCAGAACTATGCGCCGACAAATAGCGATTAACCTTCAGAACAACTGGGTAC 717
 Db 689 -----ATTGAATGGGAAGCAGGAGATCTGCTCTTCAAGTTCTAGTTATATATGC 740
 QY 718 TTCTGTAGCAATCCAGGAACCTACAAATATCAACAAGCAACACACACATCAAG 777
 Db 741 AAATTCAGCTCAACCAAGATGTCAGAAACCAACATCTGCTCCTCACTCACTTATCA 800
 QY 778 TCAAGCAAGTCAAAATATGACATGATAGTCTTGAAGACGCTCTCAAACTGCTCTT 837
 Db 801 TCA---AAATCAGGGGAAAACATTTCAGCTTTTACGTGAATTTGTAATGCTAAACCTT 857
 QY 838 GAGTCACGACATGTAGATCTGATGCGCTTGTCTTTGATCCAGCAACAATCAAGTCG 897
 Db 858 ATCAGAACGCGATGGAATCTGATGCGCTTATTTTCGACCCAGGCAAAATCAAGTCG 917
 QY 898 AACAGTGAAGTGTGAGAGTCCACAGGAGATCATTAACCACTTACCTTACTCA 957
 Db 918 AACCGCAGAGGTGATGCTGCTCATGTAACCATTAACCTTATCTTATGAACA 977
 QY 958 AATGCTGAATTTGGAAGACGATGCTGTAATTTCCCTTGTATGCTTCAACCA 1017
 Db 978 AATGCTGAATTTGGAAGACGATGCTGTAATTTCCCTTGTATGCTTCAACCA 1037
 QY 1018 TTGGGTACAGATTCAGGCGCAACCAAGTCCACACCGACTCCGGAACCTAGTCC 1077
 Db 1038 TTGGGTACAGATTCAGGCGCAACCAAGTCCACACCGACTCCGGAACCTAGTCC 1097
 QY 1078 AGGCCCCGAACCTGACCAATCTTAAATAGACTCAATTCCTC-----TTT 1125
 Db 1098 AAGTCCGCAACCTGACCAATCTTCAACCGACTCCAGCAATTCATTTGATGAAATTT 1157
 QY 1126 GGTATGAGCTGATGAGAAAGTTGGGAAGGATGATGTTTGAAGAAAGGATCTC 1185
 Db 1158 GGTCAAGAAAGCTGTTCGAAAGTGGGAGTGTATGCTTTTGAAGAAAGGATCTC 1217
 QY 1186 TCGTTATGCTTTGGGAAGATTTACATCTGAAACTGTTAAATCTTGAAGCAAGTT 1245
 Db 1218 TCGTTATGCTTTGGGAAGATTTACATCTGAAACTGTTAAATCTTGAAGCAAGTT 1277
 QY 1246 ATCAAAACAAAGAGTGTTCACACATTTAAGCTTAAAGAAATGTTGCTCTCG 1305
 Db 1278 GGCAGAAAGAGAAATTTATCTCATMACTGAGCAAGAAATGACCTCCCATCTAG 1337
 QY 1306 TGACCAAGAAATTTATGATTAAGCATATATCTGTTAAGGCTCATTAAGCTTGT 1365

Db 1338 TGATCGAAGATTTTACATAGAGCTTATGACTTACTGCAAGAAATTCACCAAGATTTACT 1397
 QY 1366 TGNAAATTAAGGTGTGTAATTTCTGATTTCCAGCTTACAGCAATTAATTAAGACCTTGA 1425
 Db 1398 TGATTAATTAAGGTGTGTAATTTCTGATTTCCAGCTTACAGCAATTAATTAAGACCTTGA 1457
 QY 1426 TGATTAATTAAGGTGTGTAATTTCTGATTTCCAGCTTACAGCAATTAATTAAGACCTTGA 1485
 Db 1458 GATATGCTCCAGATGATTAAGTCAAGTATGATGATGATGATGATGATGATGATGATGAT 1517
 QY 1486 TACCATTCAGAGAGGACTTGGCAAACTTCTCAAAATTAAGTATCTGAAGACGAGT 1545
 Db 1518 TCGTATTCAGAAAGTGTAGGAAAAACCAATTCGCAATTAACCTGATGATGATGATGAT 1577
 QY 1546 TCGTATTCAGAAAGTGTAGGAAAAACCAATTCGCAATTAACCTGATGATGATGATGATGAT 1605
 Db 1578 TCAAGTATGCAAGTGTAGGAAAAACCAATTCGCAATTAACCTGATGATGATGATGATGAT 1637
 QY 1606 TGATATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1665
 Db 1638 TGATATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1697
 QY 1666 GATTGAAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1725
 Db 1698 GATTGAAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1757
 QY 1726 AAAAGTATCTTACCTTCACTTCTTCAAGACGATGTTAAAGCAATCCAACTGAGATAG 1785
 Db 1758 GAAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1817
 QY 1786 TGACGAGCTATTTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1845
 Db 1818 AGCAAGAGCTATCTCAACCGCGTGAAGACGCTTAAGAGTGCATTTATGATGATGATGAT 1877
 QY 1846 ATATATGTTGAGATACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1905
 Db 1878 TTACATCTTCAATATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1937
 QY 1906 TCATTAATCAATATTTAAATTTGCTTGTGATGATGATGATGATGATGATGATGATGATGAT 1965
 Db 1938 CCATTAATCAATATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1997
 QY 1966 CTATACCTGGAAGATTTGTTGCAAGATTAAGTATCTGATGATGATGATGATGATGATGAT 2025
 Db 1998 GTATACCTTGAAGATTTTGTGCAAGTGTGATGATGATGATGATGATGATGATGATGAT 2057
 QY 2026 TCCACATTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2085
 Db 2058 TCCGATTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2117
 QY 2086 CAGTGAAGATTCAAATTAAGACTTCAAGCGGATGATGATGATGATGATGATGATGATGAT 2145
 Db 2118 CCAAGACAGTAAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2177
 QY 2146 TGAGCCAGAG 2156
 Db 2178 TGAATCTGATG 2188

RESULT 17
 AX568778
 LOCUS 2517 bp DNA linear PAT 29-NOV-2002
 DEFINITION Sequence 1985 from Patent WO02077021.
 ACCESSION AX568778
 VERSION AX568778.1 GI:26002455
 KEYWORDS
 SOURCE
 ORGANISM Streptococcus pneumoniae
 Streptococcus pneumoniae
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 REFERENCE 1
 Masignani, V., Tettelin, H. and Fraser, C.

TITLE Streptococcus pneumoniae proteins and nucleic acids
JOURNAL Patent: WO 02077021-A 1985 03-OCT-2002;
Chiron Spa (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)

FEATURES
source 1.2517
location/Qualifiers
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/db_xref="taxon:1313"

BASE COUNT 830 a 509 c 545 g 633 t

Query Match 41.4%; Score 990, DB 6; Length 2517;
Best Local Similarity 68.1%; Pred. No. 4,7e-201;
Matches 1478; Conservative 0; Mismatches 636; Indels 57; Gaps 5;

```

QY 1 TTCTTAGAGAGTGGAGCTGATCAAGCTGGAACGTTAAGAAA---TAATGCTTTC 57
DB 60 TTCTTAGAGAGTGGAGCTGATCAAGCTGGAACGTTAAGAAAAGCTTAATGAGCTTTC 119
QY 58 CTATATAGATGGAACCAAGCCGCAAAACGAGAAATTTGACTCTGATGAGGTTAG 117
DB 120 TTATATAGATGAGTGCAGGCTGCTCAAAAGCAGAAAATTGACACAGATGAACTCAG 179
QY 118 CAAAGCTGAAGATCAATGCTGAGCAAAATGCTCATCAAGATPAACAGCCAAAGCTATGT 177
DB 180 TAAGAGGAGGAGGATCAAGCCGCAAAATCGTCATCAAGATTAAGGATCAAGGTTATGT 239
QY 178 CACTTCACATGGGACCACTATCATATTAACAATGTAAGTTCCTTAAGAGCTATCAT 237
DB 240 GACCTTCATGAGACCATTAATCATTAATAAGGAGTCCCTTAATGATGCCATCAT 299
QY 238 CAGTGAAGATTAATCATGATAAAGATCCAAACTAATAGCTAAAGATGAGGATTTGTTAA 297
DB 300 CAGTGAAGAGCTCCTCATGAAGATCCGAATTAATCAAGTGAAGATTCAGACCTTCA 359
QY 298 TGAAGTCAAGGAGTATGTTATCAAGAGTGAAGAAATCAATGTTTACCTTAAGA 357
DB 360 TGAATCAAGGAGTATGTTATCAAGAGTGAAGAAATCAATGTTTACCTTAAGA 419
QY 358 TGCTGCCACGCGGAGTAACGTCCTGTAACAAAGAGAAATCAATCCGCAAAACAGAGA 417
DB 420 TGCAAGTCAAGGAGTATGTTATCAAGAGTGAAGAAATCAATCCGCAAAACAGAGA 479
QY 418 TAGTCAACATGCTGAAGTGAAGTCCAAAGAAAGATGAGTGTGCTTGGCAGCTTC 477
DB 480 CAGTCAATATCAAGGAGTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 533
QY 478 GGAAGGAGCTTACTACAGATGATGTTATCTTTAAATGCTTCTGATATCATAGAGA 537
DB 534 CCAAGGAGCTTACTACAGATGATGTTATCTTTAAATGCTTCTGATATCATAGAGA 593
QY 538 TACTGTATGCTTATATGTTCTTCTATGAGATCATTAACATTAATCTTAAGATA 597
DB 594 CACGGGTGATGCTTATATGTTCTTCTACGCGCACTTACATTAATCTTAAGATA 653
QY 598 GTTATCAGTAGAGAGTGTGCTGCTGAGAGGCTTCTATCGTGAAGAAATCTGTC 657
DB 654 GTTATCAGTAGAGAGTGTGCTGCTGAGAGGCTTCTATCGTGAAGAAATCTGTC 688
QY 658 AAATTCAGAACTTATGCGCGCAAAATAGCGATTAACATTCAGAAACAACTGGATAC 717
DB 689 -----ATTGGAATGGAGAGAGGATCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 740
QY 718 TTCTGTAAGCAATCCAGAACTTAACAATTAACAAGACAAACAGCACTACAG 777
DB 741 AAATTCAGAGTCAACCAAGATGTCAGAGAACCAATTCAGTGTCACTCCAACTTATCA 800
QY 778 TCAGCAAGTCAAAAGTAATGACATGATGCTCTTGAACAGCTGTACAACTGCTT 837
DB 801 TCA---AAATTCAGAGGAGAAACATTCAGAGCTTTTACGTAATTTGATGCTTAACCTT 857
QY 838 GAGTCAAGCACTGATGAATCTGATGCTTGTCTTGTGATCCAGCAAAATCAAGTGC 897

```

```

DB 858 ATCAGACGCCATGATGAAATCTGATGCGCTTATTTTCAGCCAGCGCAAAATCAAGTGC 917
QY 898 AACAGTAGAGGTGTGACAGTGCACAGGAGATGATTACCACTTATCCCTTACTCTCA 957
DB 918 AACCGCCAGAGGTGATGCTGCTTCTGATGTAACCAATTAACATTAATCTTATGAACA 977
QY 958 AATGCTGAATTTGAAGAAAGCAATGCTGATTAATTTCCCTTCTGTTATGTTCAACCA 1017
DB 978 AATGCTGAATTTGAAGAAAGCAATGCTGATTAATTTCCCTTCTGTTATGTTCAACCA 1037
QY 1018 TTGGTACCAATTTCAAGCCCAAAACCAAGTCCACACAGGACTCCGGAACCTAGTCC 1077
DB 1038 TTGGTACCAATTTCAAGCCCAAAACCAAGTCCACACAGGACTCCGGAACCTAGTCC 1097
QY 1078 AGGCGCGAAGCTGACCCCAATCTTAATAATAGCTCAAAATTCCTC-----TTT 1125
DB 1098 AAGTCGCAAGCTGACCAAAATCTCAACACTCCAAAGCAATTCATGATGAGAAAT 1157
QY 1126 GGTAGTCAAGCTGTACGAAAGTTGGGAAAGATATGATTCGAAGAAAGGCACTTC 1185
DB 1158 GGTCAAGAGAGCTGTTGCAAAAGTAGGAGTGTATGCTTTGAGGAAATGAGGTTTC 1217
QY 1186 TCGTATGCTCTTGGGAAAGATTTACATCTGAACCTGTTAAATCTTGAAGCAAGTT 1245
DB 1218 TCGTATATCCAGCCAAAGATCTTTCAGCAAGAAACAGCAGCAGGATGATAGCAAACT 1277
QY 1246 ATCAAAACAGAGAGTGTTCACACATTTAATCTCTAAAGAAAGAAAGTTGCTCTCG 1305
DB 1278 GGCACAGCAGAAAGTATATCTCATPACCTAGAGCTTAAGAAACTGACCTCCACTAG 1337
QY 1306 TGACCAAGATTTTATGATTAAGCATATATCTGTTAACTGAGGCTCAATAAGCTTGT 1365
DB 1338 TGATGAGATTTTACATTAAGCTTATGACTTATAGCAAGAAATTCACCAAGATTTACT 1397
QY 1366 TGNAAATAGAGTGTGTAATTCGATTTCCAGCCTTAAGCAAAATTAAGAGCTTGA 1425
DB 1398 TGATATTAAGAGTGCACAAGTTGATTTGAGGCTTGAATTAACCTGTTGAAGCACTCA 1457
QY 1426 TGATGATGCAATTAAGAAAGAAATGTTGATGATGATTTATGGAATCTTGCACCAT 1485
DB 1458 GATGATCCCAAGTATTAAGATCAAGTATGATGATATTTCTTGCCTTATGCTCCGAT 1517
QY 1486 TACCATCCAGAGCACTTGGCAACCAAAATCTCAAAATTTGATGATGAGAGCAAGT 1545
DB 1518 TCGTATCCAGAGCTTTAGGAAACCAATGCGAAATTAATCACTGATGATGATGAT 1577
QY 1546 TCGTATGCTCAATTAAGCTGATTAAGTAAACAGTCAAGTGTACATTTTGTATGAACA 1605
DB 1578 TCAAGTAGCAAGTTGGCAGGCAAGTACCAACAGAAAGCGTTATATCTTTGATCTCG 1637
QY 1606 TGATATTAATCAATGATGAAGAGATGCAATATTAACGCTCATATGGCCATATGTCATG 1665
DB 1638 TGATATTAACCAAGTATGAGGAGGATGCTATGTAATCAATTAAGTACCAATGACCAAG 1697
QY 1666 GATTGAAAGATAGCTCTTCTGATTAAGGAAAGTTGACAGCTCAAGCTTAATTAAGA 1725
DB 1698 GATTTAAAGATAGTGTGCTGAGAGCTGAGAGAGCGGCAAGCTTATGCTTAAGA 1757
QY 1726 AAAAGTATCTTACCTCAATCTCAAGGCAAGATTTAAAGCAATTCAACTGAGATAG 1785
DB 1758 GAAAGTATGACCCCTCTTCAAGACCATCAGATTCAGGAAATTAATGAGGCAAAAG 1817
QY 1786 TGCAGCAGTATTTACATCTGTTGTAAGGAGGAAAGCAATTCACCTGTTGCACTTCC 1845
DB 1818 AGCAAGATTAATCTCAACCGGTGAAGACACTTAAGAGGCACTGATGATGATGTC 1877
QY 1846 ATATATGTTGAGCACTGATGAGGTTAAAGAGTAAATTTGATTAATCTCATTAAGA 1905
DB 1878 TTACATCTTCAATATATCTGTAAGATCAAAAGGTAATTAATCAATCACTCATATATA 1937
QY 1906 TCATTAACATTAATTAATTTGCTGTTGATGATCAACATCAAGCTTCAAAATG 1965
DB 1938 CATTATCAATTAATTAATTTGAGTGTGATGAGGCTTTATGAGGCACTTAAGG 1997

```

QY	1966	CTATACCTTGGAAAGTTTGTTCGACACATTAACTAGTACGAGGAAACCCCTGAGAACG	2025
Db	1998	GTAATCTCTTGGAGATCTTTTGGCACTGTCAAGTACTATGTGCAACATCCAAACGAACG	2057
QY	2026	TCCACATTTCTATGATGATGATGGGCAATGCCAGTACGATGTGTGAGCAAGAAAGACCA	2085
Db	2058	TCCGATTCAGATATATGGTTTGTGTAACGCTAGCCACCATGTTCTGTAATAAATAAGTGA	2117
QY	2086	CAGTAAATCCAAATTAAGACTTCAAAGCGATGAAGCCAGTAAAGGAAACCTTGC	2145
Db	2118	CCAAACAGTAACCTGATGAAAGATAGAACATGATGAATGAGCCAACTCACCC	2177
QY	2146	TGAGCCAGAG	2156
Db	2178	TGAATCTGATG	2188
RESULT 18			
LOCUS	AE007403	11931 bp	DNA linear BCT 31-AUG-2001
DEFINITION	Streptococcus pneumoniae TIGR4	section 86 of 194 of the complete genome.	
ACCESSION	AE007403	AE005672	
VERSION	AE007403.1	GI:14972469	
KEYWORDS	Streptococcus pneumoniae TIGR4		
SOURCE	Streptococcus pneumoniae TIGR4		
ORGANISM	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.		
REFERENCE	1 (bases 1 to 11931)		
AUTHORS	Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D., Peterson,S., Heidelberg,J., Deboy,R.T., Haft,D.H., Dodson,R.J., Durkin,A.S., Weim,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D., Umayam,L.A., White,O., Salzberg,S.L., Lewis,M.R., Radune,D., Holtzapple,E., Khouit,H., Wolf,A.M., Uterback,T.R., Hansen,C.L., McDonald,L.A., Feldblum,T.V., Angiuoli,S., Dickinson,T., Hickey,E.K., Holt,I.E., Loftus,B.J., Yang,F., Smith,H.O., Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and Fraser,C.M.		
TITLE	Complete genome sequence of a virulent isolate of Streptococcus pneumoniae		
JOURNAL	Science	293 (5529), 498-506 (2001)	
MEDLINE	21357209		
PUBMED	11463916		
REFERENCE	2 (bases 1 to 11931)		
AUTHORS	Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D., Peterson,S., Heidelberg,J., Deboy,R.T., Haft,D.H., Dodson,R.J., Durkin,A.S., Weim,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D., Umayam,L.A., White,O., Salzberg,S.L., Lewis,M.R., Radune,D., Holtzapple,E., Khouit,H., Wolf,A.M., Uterback,T.R., Hansen,C.L., McDonald,L.A., Feldblum,T.V., Angiuoli,S., Gesuwan,P., Hickey,E.K., Holt,I.E., Loftus,B.J., Uwal,M.L., Yang,F., Smith,H.O., Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and Fraser,C.M.		
TITLE	Direct Submission		
JOURNAL	Submitted (29-JUN-2001)		
FEATURES	Medical Center Dr. Rockville, MD 20850, USA		
source	location/Qualifiers		
	1. 11931		
	/organism="Streptococcus pneumoniae TIGR4"		
	/mol_type="genomic DNA"		
	/strain="TIGR4"		
	/db_xref="taxon:170187"		
	129..476		
	/gene="SP0995"		
	655..795		
	/gene="SP0996"		
	/note="This region contains a match to at least one other gene that is not full length, and is not the result of a sequencing artifact; similar to GP:5019553; identified by sequence similarity; putative; IS650-spl, transposase orf2, truncation"		
	956..1156		
	gene		
	956..1156		
	5016..5933		
	gene		
	956..1156		
	/gene="SP0997"		
	/note="identified by Glimmer2; putative"		
	/codon_start=1		
	/transl_table=11		
	/product="hypothetical protein"		
	/protein_id="AAK75114.1"		
	/db_xref="GI:14972470"		
	/translation="MASASASTSTQAOQGVKSELRALSOELDRKALATVSDPK IDIKAVLDQAQPEDSALTE"		
	1303..1602		
	/gene="SP0998"		
	1303..1602		
	/gene="SP0998"		
	/note="identified by Glimmer2; putative"		
	/codon_start=1		
	/transl_table=11		
	/product="hypothetical protein"		
	/protein_id="AAK75115.1"		
	/db_xref="GI:14972471"		
	/translation="MTETNSVPAGVIVSLLALGYIAFWLIRKKESEIQSTELI KVLQDLDAEKADKKVLAQNQLDFTLDFVKEENGSAKETTKLVEELKALIDKX"		
	1833..2540		
	/gene="SP0999"		
	1833..2540		
	/gene="SP0999"		
	/note="similar to GP:6714962; identified by sequence similarity; putative"		
	/codon_start=1		
	/transl_table=11		
	/product="cytochrome c-type biogenesis protein Ccda"		
	/protein_id="AAK75116.1"		
	/db_xref="GI:14972472"		
	/translation="METIVLIVFLAVLAVLSFSPCTFLPYAGILDDQSAKSF SLTRKVLWGLRTICFLGISLIFLISFGAGYGHILYAMFRIGKAITIIGL HOMEIFLKLKLEQKSTFPKSSNRSWSPFLGIFTFEWTCTIGLVSSVALAAS GGAAWGGAIVTLIYLGMLPFLVALASGLVMPYFSKIKRHMMLKIKIGFLIVLM GILLGQVNVVLGIFB"		
	2551..3108		
	/gene="SP1000"		
	2551..3108		
	/gene="SP1000"		
	/codon_start=1		
	/transl_table=11		
	/product="thioredoxin family protein"		
	/protein_id="AAK75117.1"		
	/db_xref="GI:14972473"		
	/translation="MKRWVAGLSLISLVLMACGEETKTKTAOQPKQTTVOQIA VGDADPFTIQSGDKGVKSLSDPKKRVYLKFWASGCPKSKMPELMELAAAPDRF EILTVALPGIQSGEIVQEPQWQEOEQKDIPLVDTKATTFQAYOIRISPELYLDS OGKIKGFQFASINADAEAAFKEMN"		
	3482..4873		
	/gene="SP1001"		
	3482..4873		
	/gene="SP1001"		
	/note="identified by match to Pfam protein family HNM PF00324"		
	/codon_start=1		
	/transl_table=11		
	/product="amino acid permease family protein"		
	/protein_id="AAK75118.1"		
	/db_xref="GI:14972474"		
	/translation="NNIPRTKVNSLDKTEWHRHLKMLDILILGIGAVNGVFTITG AATTLAQRALVITIVSALCTGVLSALFPFAPARVATGAGVAYLVAIGEPFAMVIG WLTMMEPTRDISGVAAGMAVYFKGLISQGIAPQALNGTFRNQAGTFPDLFIVLV LTVSLVLMKAKALRFNSILVILKFSALAEVLVGVNWKEDWNSNAPYGFQIIGA STGIMAGSLMPFGFESISMAVDEVKTPQNIIPRGIVLSISIVTILVALVTLVLT GVAVHSHLNDADVALRSVGSIMANVSLVAIPLITVCISMTYALSMTYLSAR DGVLPAPAKELTSTKIPKATILTGASVAAAGMPLASIAAFINICTLAYIMLAY GLIRKREKMPKRGFKTPVLVPLILSLISLSTFMLQYNNMTWIAFLVALLVGSII YFTYGVKHSITTE"		


```

Db      6978 TTGGGTACAGATTCAGACGCAAGCAACCAAGTCCACATGACCTCCGAGCACTAGTCC 7037
Qy      1078 AGCCCCGCACTGCAACCAATCTTAAATAGACTCAATTTCTC-----TTT 1125
Db      7038 AAGTCGCAACCTGCAACCAATCTTCAACACCTCCCAATTCATTTGATGAGAAAT 7097
Qy      1126 GGTATGTCAGTGTAGCAAAAGTTGGGAGGATATGATTCGAGAAAGAGGATCTC 1185
Db      7098 GGTCAAGAGAGCTGTTCCAAAGTGGGCGATGTATGCTTTGGAGAGATGAGTTTC 7157
Qy      1186 TCGTTATGCTTTGGCAAGATTTACATCTGAACTGTTTAAATCTTGAAGCAAGT 1245
Db      7158 TCGTTATATCCAGCAAGAGTCTTTCAGCAGAAAGAGCAGCAGCATTTGATGAACT 7217
Qy      1246 ATCAAAACAGAGAGTGTTCACACACTTAACTGCTAAAGAAAGAAATGTTGCTCTCG 1305
Db      7218 GCCCAAGCAGGAAATTTATCTCATAGCTAGAGCTTAAAGAACTGACCTCCATCTAG 7277
Qy      1306 TGACCAAGATTTTATGATTAAGCATATATCTGTTAACTGAGGCTTAAGGCTTGT 1365
Db      7278 TGATGAGATTTTACATTAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACT 7337
Qy      1356 TGNAAATAAGGCTCGTAAATTCGATTTCCAGCCTTAGACAAATTTAGAACGCTTGA 1425
Db      7338 TGATATATAAGGTCGACAAAGTGTGATTTTGAGGCTTGGATACCTGTTGAAGCACTCA 7397
Qy      1426 TGATATATGACTAATAAGAAAAATGGTAGATGATTTATGCGATCTTTCAGCAAT 1485
Db      7398 GGATGTCCTCAAGTGTAAAGTCAAGTGTGATATATCTTGGCTTTTAGCTCCAT 7457
Qy      1486 TACCATCCAGAGCACTTGGCAACCAATCTCAATTTGATGATGATGAGAGCAAGT 1545
Db      7458 TCGTATCCAGAGCTTTAGGAAACCAATCCGAATTTACTTCACTGATGATGAAAT 7517
Qy      1546 TCGATTGCTCAATTAGCTGATTAAGTATACACGTCAGATGCTTACATTTTGTATGAACA 1605
Db      7518 TCAAGTACCAAGTTGGCAGGCAAGTACCAACAGACGGTTATATCTTTGATCTCG 7577
Qy      1606 TGATATATCAAGTATGAAGAGATGATATGTAACGCTCATATGGGCGATGACTG 1665
Db      7578 TGATATTAACCAAGTATGAGGGGAGTGCCTATGTAACCTCAATATGACCCATAGCACCTG 7637
Qy      1666 GATTGAAAGATGATCCCTTTGATTAAGGAAAAATGGAGCTCAAGCTTACTAAAGA 1725
Db      7638 GATTAAAAAGATGATTTGTCTGAACTGAGAGCGGAGCCGAGCTTATGCTTAAAGA 7697
Qy      1726 AAAAGTATCTTACCTCATCTCCAGAGCAGATGTTAAAGCAATCCCACTGAGATAG 1785
Db      7698 GAAAGTTTGACCCCTCTTGACAGACCATCAGGATTCAGGAATATCGAGGCAAAAG 7757
Qy      1786 TGCACAGCTATTTTCAATCGTGTGAAAAGGGAAGAAAGCAATTCACCTGTTGCACTCC 1845
Db      7758 AGCAGAACTATCTTCAACCGGCTGAAAAGAGCTAAGAAAGGCGCACTTGATCGTATCC 7817
Qy      1846 ATATATGTTGAGCATCAGTGAAGTAAAAAGGTAATTTGATTATTCCTCATAAAGA 1905
Db      7818 TTACATCTTTCATATCTGTAAGTCAAAAGCGTAGTTTATCATATCTCATTAAGA 7877
Qy      1906 TCATTACATATATTAATTTGCTTGGTTGATGATCACACATACAAAGCTCCAAATAG 1965
Db      7878 CCATTACATATATCAATTTGAGTGTGTTGACAAAGGCTTTATGAGGCACTTAAGG 7937
Qy      1966 CTATATCTTGAAGATTTGTTGGCAGATTAAGTACTAGTAAAGCAACCTGAGCAAG 2025
Db      7938 GTATCTCTTGGAGATCTTTTGGCGATGTCAGTCAAGTACTATGCAACATCCAAAGCAAG 7997
Qy      2026 TCCACATCTTATATGATGAGGGAATGCGCATGAGATGTTTGAAGCAAGAAAGACA 2085
Db      7998 TCCGATTCAGATATATGTTTGGTGAAGCTAGCAGCATGTTGTTAAATTAAGTAGA 8057
Qy      2086 CAGTGAAGATCCAAATAGAACTTCAAGCGGATGAAGCCAGTGAAGAAACACCTGC 2145

```

```

Db      8058 CCAAGACAGTAACCTGATGAGATTAAGSAAATGATGATGAGGCCACTCACCC 8117
Qy      2146 TGAGCCAGAG 2156
Db      8118 TGAATCTGATG 8128

RESULT 19
AX571762
LOCUS      AX571762      349980 bp      DNA      linear      PAT 29-NOV-2002
DEFINITION Sequence 4981 from Patent WO02077021.
ACCESSION AX571762
VERSION   AX571762.1 GI:26003954
KEYWORDS
SOURCE
ORGANISM  Streptococcus pneumoniae
           Streptococcus pneumoniae
           Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
           Streptococcus.
REFERENCE 1
AUTHORS   Maeligani,V., Tettelin,H. and Fraser,C.
TITLE      Streptococcus pneumoniae proteins and nucleic acids
JOURNAL    Patent: WO 02077021-A 4981 03-OCT-2002.
           Chiron Spa (IT) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
           source
           1. 349980
              /organism="Streptococcus pneumoniae"
              /mol_type="genomic DNA"
              /db_xref="taxon:1313"
              /note="seq 4979 too long: 2,162,598 bases-replaced by
              following seq:-seq 4979; from 0.000,001 to 0.349,980-seq
              4980: from 0.300,001 to 0.649,980-seq 4981: from 0.600,001
              to 0.949,980-seq 4982: from 0.900,001 to 1.249,980-seq
              4983: from 1.200,001 to 1.549,980-seq 4984: from 1.500,001
              to 1.849,980-seq 4985: from 1.800,001 to 2.149,980-seq
              4986: from 2.100,001 to 2.162,598"
BASE COUNT 108592 a 61526 c 75542 g 104320 t
ORIGIN
Query Match      41.4%; Score 990; DB 6; Length 349980;
Best Local Similarity 68.1%; Pred. No. 4,26-201;
Matches 1478; Conservative 0; Mismatches 636; Indels 57; Gaps 5;

Qy      1 TTCTTACGAGTGGGAGCTGTATCAAGCTAGAACGGTTAAGGAAAA--TAACTGTTTTC 57
Db      238008 TTCTTATGAAGACTTGTCGTGTCACCAAGCTGTGCTGATTAAGAAAGCTATTCAGTTTC 238067
Qy      58 CTATATATGATGAAAAACAGCGACGCAAAAAACGAGAAATTTGACTCTTGATGAGTTAG 117
Db      238068 TTATATATGATGTGATCAGCTGTGTCAAAAAGCGAAAAAATTGACACAGATGAAGTCAG 238127
Qy      118 CAAGCTGAAAGATTCATATGCTGAGCAAAATGTCATCAATGAATACAGCAAGGCTATGCT 177
Db      238128 TAAAGGGAGGGAGTCAACCGCAAAATGTCATCAATGAATTAAGATCAAGGATTAATGT 238187
Qy      178 CACTTCATGAGCGACCACTATCATTTATCAATGTAAGTTCCTTATGACGCTATCAT 237
Db      238188 GACCTCTCATGAGACCATATATCTATTAATGCAAGAGTCCCTTATATGTCATCAT 238247
Qy      238 CAGTGAAGAAATTAATCAATGAAGATCCAAACTATTAAGCTTAAAGATGAGATATTTGTTAA 297
Db      238248 CAGTGAAGAGCTCCTCATGAAGATCCGAATTTACATTAAGAGATTCAAGCATTTGTCAA 238307
Qy      298 TGAGGTCAAGGTCGATATGTTTATCAAGTATGATGAAATTAATGTTTAACTTAACTTAAGGA 357
Db      238308 TGAATCAAGGTCGTTATGTTTCAAGGTAGATGAAATTAATCATATTTTAACTTAAAGGA 238367
Qy      358 TGCTGCCAGCGGATTAACGTCGTAACAAAGAGAAATCAATGACAAAGAAACAGAGCA 417
Db      238368 TGCAGCTCATGCGGATTAATTTGGACAAAAGAAAGAAATTAAGCTGAGAGAGAAACA 238427
Qy      418 TAGTCAACATCTGTGAAGTGAAGCTCCAGAAAGATGATGCTGTTGCTTGGCAGCTTC 477
Db      238428 CAGTCAATATCAAGGCGGTGTTCT-----AACGATCAAGCAGTATGTCAGACCAAGC 238481

```

QY 478 GCAAGAGCGCTACTACATGATGATGTTATCTTTAATGCTTCATATCATAGAGA 537
 Db 238482 CCAAGAGCGCTACTACATGATGATGTTATCTTTAATGCTTCATATCATATGAGAGA 238541
 QY 538 TACTGTGATGCTTATATCTTCTTCATGAGATCATTAACATTCATCTTAAAGATGA 597
 Db 238542 CACGGGTGATGCTTATATCTTCTTCATGAGATCATTAACATTCATCTTAAAGATGA 238601
 QY 598 GTTATCAGTACGAGTGGCGTCGACAAAGCCTTCATCTGTCGAGGAAATCTGTC 657
 Db 238602 GTTATCAGTACGAGTGGCGTCGACAAAGCCTTCATCTGTCGAGGAAATCTGTC 238636
 QY 658 AAATTCAGAAACCTATCGCCGACAAAATAGCATTAACCTTCAAGAACAACTGGGTAC 717
 Db 238637 -----ATTGGAAATGGAGACAGGATCTGCTCTTCTTCAAGTTCAGTTATTAATGC 238688
 QY 718 TTCTGTAGCAATCCAGAACTACAAATATCTAACACAGCAACACAGCACTAACAG 777
 Db 238689 AAATCCAGCTCAACCAAGATGTGACAGAAACCAATCTGATCTCATCTCAACTTATCA 238748
 QY 778 TCAAGCAATCAAGATATGACATGATATGCTCTTGAACAGCTCTACAACTGCTTT 837
 Db 238749 TCA-----AAATCAAGGGGAAAAATTCAGCCCTTTTACGTGAATGTATGCTAAACCTTT 238805
 QY 838 GAGTCAGAGATGTGAGATCTGATGAGCCTTGTCTTGTATCCAGACAAATCAAGTGC 897
 Db 238806 ATCAGAGCGCATGTGAGATCTGATGAGCCTTGTCTTGTATCCAGACAAATCAAGTGC 238865
 QY 898 AACAGCTAGAGGTGTGAGTGCACACAGAGATCATTAACATTCATCTTCTTACTCTCA 957
 Db 238866 AACCCGAGAGGTGTGAGTGCACACAGAGATCATTAACATTCATCTTCTTACTCTCA 238925
 QY 958 AATGCTGAATGTGAAGAACGATGCTGCTGATTTATCTCCCTTCTGTTATCTTCAACCA 1017
 Db 238926 AATGCTGAATGTGAAGAACGATGCTGCTGATTTATCTCCCTTCTGTTATCTTCAACCA 238985
 QY 1018 TTGGGTACAGATTCAGAGCGCAGAACCAACAGTGCACAAACGAGCTCCGGAACCTAGTCC 1077
 Db 238986 TTGGGTACAGATTCAGAGCGCAGAACCAACAGTGCACAAACGAGCTCCGGAACCTAGTCC 239045
 QY 1078 AGGCCGCAACCTGCAGCAAAATCTTAAATAGACTCAAAATCTTCTC-----TTT 1125
 Db 239046 AAGTCGCAACCTGCAGCAAAATCTTCAACAGCTCCAAACATTCATTCATGATGAGAAAT 239105
 QY 1126 GGTAGTCACTGCTGACGAAAGTTGGGAGAGATGTATTCGAAGAAAGGCACTCTC 1185
 Db 239106 GGTCAAGAGCTGTGCAAGAGTGGCGATGTATGTCTTGAAGAAAGGCACTCTC 239165
 QY 1186 TCGTATATCTTGGGAAAGATTTACATCTGAACCTGTTAAATCTTGAAGAGCAAGT 1245
 Db 239166 TCGTATATCTTGGGAAAGATTTACATCTGAAGCTTAAGAGAGCAAGT 239225
 QY 1246 ATCAAAACAGAGAGTGTTCACACATTTAACTGCTAAAGAAAGAAATGTGCTCTCG 1305
 Db 239226 GGCACAGAGAGAGTGTTCACACATTTAACTGCTAAAGAAAGAAATGTGCTCTCG 239285
 QY 1306 TGACCAAGATTTTATGATTAAGCATATATCTGTTAAGTGAAGCTCATTAAGCTTGT 1365
 Db 239286 TGATGAGAGATTTTATGATTAAGCATATATCTGTTAAGTGAAGCTCATTAAGCTTGT 239345
 QY 1366 TGNAAATAAGGTCGTAATCTGATTTCCAGACCTTAGCAAAATTAATGAAGCTTGA 1425
 Db 239346 TGATTAATAAGTTCGTAATCTGATTTTGAAGCTTAGCAAAATTAATGAAGCTTGA 239405
 QY 1426 TGATGAATGCTAATTAAGAAATAATGCTAGATGATTTATGCAATTCCTAGCACAAT 1485
 Db 239406 GGATGTCCTCAAGTGAATTAAGTCAAGTATGATGATTTATCTGCTTCTTATGCTCCAT 239465
 QY 1486 TACCATTCAGAGCGAGCTTGGCAACCAATTCCTCAATTTGATGATCTGAAGAGCAAGT 1545
 Db 239466 TCGTATCCAGAACTGTTAGGAAACCAATTCGCAAAATTAACATGATGATGATGAT 239525

QY 1546 TCGATATCTCAATTAAGCTGATTAAGTATACAAAGTCAAGTGTATCTTTTATGATGAACA 1605
 Db 239526 TCAAGTACCAAGTGTGGAGGCAAGTACAAACAGAAACGTTATATCTTTGATCTCG 239585
 QY 1606 TGATTAATCAAGTATGAGAGATGATATGTAACGCTCATATGAGGCAATGACTCG 1665
 Db 239586 TGATTAATCAAGTATGAGAGATGATATGTAACGCTCATATGAGGCAATGACTCG 239645
 QY 1666 GATTGGAAGATAGCTCTTCTGATTAAGAAAGATGCTCAAGCTTATCTTAAAGA 1725
 Db 239646 GATTGGAAGATAGCTCTTCTGATTAAGAAAGATGCTCAAGCTTATCTTAAAGA 239705
 QY 1726 AAAAGTATCTTACCTCCATCTCCAGACGAGATGTTAAACCAATCCAACTGGAGATAG 1785
 Db 239706 GAAAGTGTGACCCCTCTCTTGCACAGACATCCAGATTCAGAAATATCTGAGGCAAAAG 239765
 QY 1786 TGACGAGCTTATTAATCAATCGTGTGAAGGGGAAAAAGCAATTCACCTGTTGCACTTC 1845
 Db 239766 AGCAGAACTATTAACAACCCGCTGAAGAGCACTTAAGAAAGGTGCCACTTGTATCTGCC 239825
 QY 1846 ATATATGTTGAGATATACATGATGAGTTAAACGGTAAATTTGATATCTCTCATAGAGA 1905
 Db 239826 TTACATCTTGAATATATTAATTTGCTTGTGATGATCACACATPACAAAGCTCCAAATGG 239885
 QY 1906 TCATTAACATATATTAATTTGCTTGTGATGATCACACATPACAAAGCTCCAAATGG 1965
 Db 239886 CCATTAACATATATTAATTTGCTTGTGATGATCACACATPACAAAGCTCCAAATGG 239945
 QY 1966 CTATACCTTGAAGATTTGTTGCGACGATTAAGTACTAGTGAACAACCTGACGAAG 2025
 Db 239946 GTATACCTTGAAGATTTGTTGCGACGATTAAGTACTAGTGAACAACCTGACGAAG 240005
 QY 2026 TCCCATTTCTATATGATGATGAGGCAATGCCAGTACATGTGTTAGGCAAGAAAGACCA 2085
 Db 240006 TCCCATTTCTATATGATGATGAGGCAATGCCAGTACATGTGTTAGGCAAGAAAGACCA 240065
 QY 2086 CAGTGAATCCAAATAAGAACTTCAAGCGGATGAAGAGCAGTAGAGAAACCTGTC 2145
 Db 240066 CCAAGACAGTAAACCTGATGAAGATGAAGAAACATGATGAAGTGAAGCAACTCACC 240125
 QY 2146 TGAGCCAGAG 2156
 Db 240126 TGAATCTGATG 240136
 RESULT 20
 SPNEU1901 232807 bp DNA linear HTG 12-JUL-2001
 LOCUS Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***
 DEFINITION AL449923
 ACCESSION AL449923.1 GI:11545148
 VERSION HTG; HTGS_PHASE2.
 KEYWORDS Streptococcus pneumoniae
 SOURCE Streptococcus pneumoniae
 ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
 REFERENCE 1
 AUTHORS Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Humbert, Y., Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de Francesco, M., Polissi, A., Buell, G., Feger, G., Garcia, E., Peitsch, M., and Garcia-Bustos, J. F.
 TITLE Annotated draft genomic sequence from a Streptococcus pneumoniae type 19F clinical isolate
 JOURNAL Microb. Drug Resist. 7 (2), 99-125 (2001)
 MEDLINE 21335329
 PUBMED 11442348
 REFERENCE 2 (bases 1 to 232807)
 AUTHORS Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A., Humbert, Y., Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de Francesco, M., Buell, G., Feger, G., Garcia, E., Peitsch, M., and Garcia-Bustos, J. F.
 TITLE Direct Submission
 JOURNAL Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,

Severo Ochoa 2, 28760 Tres Cantos, SPAIN
 * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.

FEATURES

Location/Qualifiers

1..232807

/organism="Streptococcus pneumoniae"

/mol_type="genomic DNA"

/serotype="19F"

/db_xref="taxon:1313"

/clone="G54"

BASE COUNT 70984 a 43538 c 47937 g 70253 t 95 others
 ORIGIN

Query Match 41.4%; Score 988.4; DB 2; Length 232807;
 Best Local Similarity 65.6%; Pred. No. 9.4e-201;
 Matches 1616; Conservative 0; Mismatches 697; Indels 150; Gaps 6;

1 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAAGGTTAAGGAAA--TAACTCGTGTTC 57
 200352 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAAGGTTAAGGAAA--TAACTCGTGTTC 57
 58 CTATATAGATGGAACAAAGCGCAAAAGGCGAAGATTGACTCTGATGAGTTAG 117
 200412 TTATATAGATGATGATCAGGCTGTCAAAAGCGCAAAAGTTGACACAGATGAAGTCAG 200471
 118 CAAGGTGAAGGAATCAATGCTGAGCAAAATGCTCAAGATTAACAGCAAGGCTATCT 177
 200472 TAAAGGGAAGGAGATCAACGCGCAAAATGTTATCAAGATTAAGATCAAGGTTATGT 200531
 178 CACTTCACATGCGCAGCACTATCATTTATTAACATGTAAGGTTCTTATGACGCTATCAT 237
 200532 GACCTCTATGAGAGCATTTATCATTTATGCAAGGTTCTTATGATGACATCAT 200591
 238 CAGTGAAGAATTACTCATGAAAGATCCAAACTATTAAGCTAAAGATGAAGATTTGTTAA 297
 200592 CAGTGAAGAAGCTCCCATGAAAGATCCCAATTAATCAATTGAAGATTCAGACATTTGCAA 200651
 298 TGAAGTCAAGGAGTGAATGTTATCAAGGTAGTGAAGATTAATGTTTACCTTAAAGA 357
 200652 TGAATTCAGAGGAGTGAATGTTATCAAGGTAGTGAAGATTAATGTTTACCTTAAAGA 200711
 358 TGCTGCCACCGCGATAAGTCCGTAACAAAGAGGAATCAATCGACAAAACAAAGACA 417
 200712 TGCAAGCTCATCGGATTAATTCGACCAAAAGAAAGATTAACGTCGAAGCAGAGACG 200771
 418 TAGTCAACATCGTGAAGGTGAATTCGAAGAAAGATGCTGTGCTTGGCAGCTTC 477
 200772 CAGTCATATCAT-----AACTCAAGAGCAATATGCTGTGCTGACGACAGAC 200822
 478 GCAAGGAGCTATATACATGATGATGATTAATCTTTAAGTCTTGATATCATAGAGA 537
 200823 CCAAGAGAGCTATACACAGATGATGATTAATCTTTAAGTCTTGATATCATAGAGA 200882
 538 TACTGTGATGCTTATGCTGCTCATGAGATGATTAACATTAATCTTCCATGAAGTGA 597
 200883 TACTGTGATGCTTATGCTGCTCATGAGATGATTAACATTAATCTTCCATGAAGTGA 200942
 598 GTTATCACTGACGAGTGGCTGCTGCAAGACCTTCTTATCTGTGAGAGAAATCTGTC 657
 200943 TTTATCTCCAGTGAATTAGCTGCGCCAGACCTATTGGAATG----- 200987
 658 AAATTCAGAAACCTATCGCCGACAAATAGCATTAACACTTCAGAGAAACAACTGGGTAC 717
 200988 -GAAGCAGAGATCTGCTCTTCTTCAAGTTCATGATTAATGCAATCAGCTCAACCA 201046
 718 TTCTGTAGCAATCAGAGAACTAATAATTAACAAGCAACAGCAACACATCAAG 777
 201047 GATTCAGAGAAACCAATCT-----GACTGTCATCTCCAACTTA 201086
 778 TCAAGCAAGTCAAGATTAATGATTAATGATGCTCTTGAAGACGCTTCAAACTGCTTT 837

201087 TCATCAAAATCAAGGGGAAAAATTTCAAGCTTTTACGTGAATGTATGCTAAACCTT 201146
 838 GAGTCAACGACATGTAGATCTGATGAGCTTGTCTTGTATGCAGCAAAATCACAAGTCG 897
 201147 ATCGAAACGCAATGTGAATCTGATGAGCTTATTTTGCAGCCAGCGAAATCACAAGTCG 201206
 898 AACAGTAGAGGTGTTGACGTGCGACACGAGATCATTTACACTTCACTTCTCTCTCA 957
 201207 AACCGCAGAGGTGTACTGCTGCTTCAAGGTAACATTAACACTTATTCCTTATGAACA 201266
 958 AATGTCGAATTTGGAAGAAGATCGCTGATTTATTCCTTCTGTTATGTTTAAACCA 1017
 201267 AATGTCGAATTTGGAAGAAGATCGCTGATTTATTCCTTCTGTTATGTTTAAACCA 201326
 1018 TTGGGTACAGATTCAGAGCGACAGAACCAAGTCACCAACCGACTCCGGAACCTAGTCC 1077
 201327 TTGGGTACAGATTCAGAGCGACAGAACCAAGTCACCAACCGACTCCGGAACCTAGTCC 201386
 1078 AGGCCGCAACCTGACACCAATCTTAAATAGACTCA-----ATTCTTCTTT 1125
 201387 AAGTCCGCAACCTGACACCAATCTTCAACAGCTCCAGCAATTCATGTATGAGAATTT 201446
 1126 GGTATGACAGTGTACGAAAGTGGGGAAGATATGTATGGAAGAAAGGCATCTC 1185
 201447 GGTCAAAAGAGCTGTGGAAGAGTGAAGCATGTATGTCTTTGAGAGAGATGAGTTTC 201506
 1186 TCGTTATGCTTTTGCAGAAAGTTTACCATCTGAAACCTGTTAAATACTTGAAGCAAGTT 1245
 201507 TCGTTATGCTTTTGCAGAAAGTTTACCATCTGAAACCTGTTAAATACTTGAAGCAAGTT 201566
 1246 ATCAAAACAAAGAGTGTTCACACACTTAACTGCTTAAAGAAAGAAATGTTGCTCTCG 1305
 201567 GGCACACAGAAAGTTTATCTCATAGCTAGAGCTTAAGAAACCTGACCTCCATCTAG 201626
 1306 TGACCAAGATTTTATATTAAGATTAATCTGTTAACTGAGCTGATTAAGCTTGT 1365
 201627 TGATCGAAGATTTTAAATTAAGCTTAACTGATTAAGCTGATTAAGCTTGT 201686
 1366 TGNAAATTAAGGAGTGTATCTGATTTCCAAAGCTTGAACAAATTAATTAAGACCTGAA 1425
 201687 TGATTAATTAAGGATGACAAAGTATTTGAGCTTTGAGATACCTGTTGGAACGACTCA 201746
 1426 TGATGAATGCACTAATTAAGAAATTTGATGATTTATGTCATTTCTAGCAACAT 1485
 201747 -GATGTCCTCAAGTATTAAGTCAAGTATGAGAGATATTTCTGCTTACGCTCCGAT 201806
 1486 TACCATTCAGAGGACTTGGCAACCAATTTCCAAATTAAGTATGATTAAGAGCAAGT 1545
 201807 TCGTCATCCAGAAAGTGTAGAAAACCAATTCGCAAAATTAACCTACACTGATGATGAGAT 201866
 1546 TCGTATTCCTCAATTTAGCTGATAGTATTAACAGCTCAGATGTTATTTTGAATGAACA 1605
 201867 TCAAGTACCAAGTTGGCAGAGCAATGACAAACGAAAGACGTTATTTCTTGATCTCG 201926
 1606 TGATATTAATCAATGATAGAGAGATGATTAAGTAAACGCTCATATGAGGCAATGACTG 1665
 201927 TGATATTAACAGATGATAGAGGAGATGCTTATGTAATTCACATATGACCATACCACTG 201986
 1666 GATTGGAAGAAATGCTTTCTGATTAAGAAAGTTGCAAGCTTAACCTTATTAAGA 1725
 201987 GATTGGAAGAAATGCTTTCTGATTAAGAAAGTTGCAAGCTTAACCTTATTAAGA 202046
 1726 AAAAGTATCCTACCTCATCTCCAGAGCGAGATGTTAAAGCAATCCAACTGAGAGATAG 1785
 202047 GAAAGTATTAACCTCTCTCTGACAGACATCAGAGATTCAGAGAAATCTGAGGCAAAAG 202106
 1786 TGACAGACTATTTAATCAATCGTGTGAAGGGAAGAAAGCAATTCACACTGTTGCACTTC 1845
 202107 AGCAGAGATCTATTAACAACCGCGTGAAGCAGCTTAAGAGGTGCACTTATGTCATGCTC 202166
 1846 ATATATGCTTGAGATTAAGTTGAGTTAAAGAAAGGTTAATTTGATTTCTCTTAAGA 1905
 202167 TTACAAATCTTCAATATATCTGATGAGATCAAAACGATGTTTATATATCTTATCTTATGA 202226

QY	1906	CCATTACCCCTATATATTAATTTGCTTGTTGATGATCCACATACAAAGCTCCAAATGG	1965
Db	202227	CCATTACCCATTAATTAATTTGAGTGGTTTGAACGAAAGCTTTATAGAGCACTTAAGGG	20220866
QY	1966	CTATACTTGAAGATTTGTTGCGACGATTAAGTACTACGTAAACACCTGACGCAACG	2025
Db	202287	GTAATCTTTGAGGATCTTTTGCGACTGTCGAAAGTACTATGTGGAACATCCAAACGAAAG	2023466
QY	2026	TCCACATTTCTAATGATGATGGGCAATGCTCACTGACATGTGTGTTAGCAAGAAAGCCA	2085
Db	202347	TCCGATTTCAAGTAAAGGTTTGGTAAACGCTACCGACCATGTTCAAAGAAACAAAATGG	2024086
QY	2086	CAGTGAAGATCCAAATPAAGAACTTCAAAGCGATGTAAGG-----	2125
Db	202407	TCAAGCTGATACCAATCAACGGAACAAACGAGAGAAACCTCAGACAGAAAAACC	2024686
QY	2126	-----	2125
Db	202467	TGAGGAAGAAACCCCTCGAAGAAAGAAACCGCAAGACGAAACCAAGTCTCCAAACC	2025286
QY	2126	-----CCAGTGAAGAAACACCTGCTGAGCCAGAAAGTCCCTCAAGTACGACTGA	2175
Db	202527	AACAGACGAACCGAAGAAAGAAATCACCAAGAGATACAGAAACCTCAGGTCCAGACTGA	2025866
QY	2176	AAAAGTAGAAGCCCAACTCAAAAGAACGAGAAAGTTTGGCTTGCGAAAGTAAAGGATCTAG	2235
Db	202587	AAAGTTGAAGAAAAAATCTAGAGAGGCTGAAGATTTACTTTGAAAAAATCCAGATCCAAAT	2026486
QY	2236	TCTGAAAGCAATGCACAGAAACTCTGACGTGTTTACGAAATATTTTGAATCTCTCAAT	2295
Db	202647	TATCAAGTCCAAATGCCAAAGAGACTCTCAGAGATTAAAAAATTAATTTACTATTTGGCAC	2027066
QY	2296	TATGATTAACATAGATCATGCGAAGACAGAAAAATTAATCTTGGCTTGTAAAGAGAG	2355
Db	202707	CCAGAGCAACAACTACTATTATATGCGACAGAACTGAAAAAATAATTTGCTTTATTAAGAGAG	2027666
QY	2356	TAA 2358	
Db	202767	TAA 202769	
RESULT 21			
LOCUS	ARI120270	2290 bp	DNA
DEFINITION	Sequence 65 from patent US 6159469.		linear
ACCESSION	ARI20270		
VERSION	ARI20270.1		
KEYWORDS	GI:14103846		
SOURCE	Unknown.		
ORGANISM	Unknown.		
REFERENCE	Unclassified.		
AUTHORS	1 (bases 1 to 2290)		
TITLE	Choi, G.H., Kunsch, C.A., Barash, S.C., Dillon, P.J., Dougherty, B.,		
JOURNAL	Fannon, M.R. and Rosen, C.A.		
FEATURES	Streptococcus pneumoniae antigens and vaccines		
source	Patent: US 6159469-A 65 12-DEC-2000;		
	location/Qualifiers		
	1..2290		
	/organism="unknown"		
BASE COUNT	766 a 474 c 498 g 547 t		5 others
ORIGIN			
Query Match	41.3%;	Score 987.6;	DB 6; Length 2290;
Best Local Similarity	67.7%;	Pred. No. 1.5e-200;	
Matches 1481;	Conservative 0;	Mismatches 645;	Indels 60; Gaps 5;
QY	1	TTCTTACGAGTTGGAGCTGTATCAAGCTGAACGTTAAGAAAA--TAATCGTGTTC	57
Db	4	TTCTTATGAACTTGGTCGTACCAACGCTGTCAGTTAAGAAAGTCTAATCGAGTTTC	63
QY	58	CTATATAGTGAAGAAACGACGCAAGAAAAACGAGAAATTTGATCTCGATAGAGTTAG	117

Db	64	TTATATGATGATGATCAGGCTGGTCAAAAAGCGAGAAAACCTTGACACCAATGAAATGAC	123
Qy	118	CAAGCGTGAAGAAATCAATGCTGAGCAAAATCGTCATCAAGATTAACAGACCAAGCTATGT	177
Db	124	TTAAGAGGAGGGGATCAACGCCGAACAAATNGTATCAAGATTAACGATCAAGGTATATGT	183
Qy	178	CACCTTCAATGGGACACCATCATCTTATTTAACAATGTAAGTTCCTTATGACGATATCAT	237
Db	184	GACCTTCAATGAGAACCATTAATCTTATCTTAAATGCGAAGTTCTTATGATGCAATCAT	243
Qy	238	CAGTGAAGAAATTACTCATGAAAGATCCAAACTAATAGCTAAAGATAGATATTTGTAA	297
Db	244	CAGTGAAGAGTCTCTCATGAAGAATCCGAATTAATCATTTGAAGATTCAAGCATTTGTCA	303
Qy	298	TGAGCTCAAGGGGTGATATGTTATCAAGTATGATGAGAAATACTATGTGTTACTTTAAGA	357
Db	304	TGAATCAAGGGTGTATGTCAATTAAGTAAACGTAATAATCATATGTTTACCTTAAGA	363
Qy	358	TGCTGCCACGCGGATTAACGTCGCTAATAAAAAGAGAAATCAATGACAAAAACAAGACA	417
Db	364	TGACGCTCATGCGGATTAATTTGCGACAAAAGAAATTAACGTACGAAGACGAGAACG	423
Qy	418	TAGTCAACATCGTAAAGGTGAAATCCCAAGAAACGATGAGTCTTGCTTGGACAGTTC	477
Db	424	CAGTCATATATCAT-----ACTCAAGACGAGATTAATCTGTCTGACGCAAGC	474
Qy	478	GCAAGACGCTATATCTACAGATGATGTTATTTATTTATGCTTCTGATATCATAGAAGA	537
Db	475	CCAAGACGTTATATCAACGAGATGATGAGTATATCTTCAATGCAATCTGATATCATTTAGA	534
Qy	538	TACTGTGATGCTTATATCGTTCTCATGAGATCATTTACATTAACATCTTAAGATGA	597
Db	535	CACGGGATGCTTATATCGTTCTCATCGCGACCATTAACATTAACATCTTAAGATGA	594
Qy	598	GTATTCAGCTAGCAGATTTGGCTGCTGAGAGACCTTCCTATCTGTGAGAGAAATCTGTC	657
Db	595	GTTATTCAGCTTAGCGAGTTAGCTGCTGAGAGCTT-----629	629
Qy	658	AAATTCAGAACTTATGCCGACCAAAATAGCGATTAACCTTCAGAAACAACTGGTAC	717
Db	630	-----ATTGGAATGGAGAGAGGATCTCGTCTTCTTCAAGTTCTAGTTAATATGC	681
Qy	718	TTCTGTATGCAATTCAGGAATCTACAAATCTTAACACAGCAACAACGCAACTTAACAG	777
Db	682	AAATTCAGCTCAACCAAGATTTGTCAGAAACCAACATCTGACTCACTCAACTTTATCA	741
Qy	778	TCAGCAAGTCAAAAGTATGATCATTTGATGATGCTTGAACAGCTCTACAAACGCTTT	837
Db	742	TCA--AAATTCAGGGGAAAAACATTTCAAGCTTTTACGTGAATTTGATGTAAACCTT	798
Qy	838	GAGTCAACGACATGTAAATCTGATGCGCTTGTCTTTGATCTCAGCACAATACTCAATGCG	897
Db	799	ATCGAAGCGCATGTGAAATCTGATGCGCTTATTTTGACCCAGCGCAAAATCAAAATGCG	858
Qy	898	AACAGTAAAGGTGTTCAAGTGCACACGAGATCATTTACACTTCAATCCCTTACTCTCA	957
Db	859	AAACGACAGAGGTGATCTGCTCTCTCATGTAACTTAACACTTATCCCTTATGAAACA	918
Qy	958	AATGTCGAATTTGGAAGAAAGAAATCGTCTGATATATCCCTTGGTATATGTTAAACA	1017
Db	919	AATGTCGAATTTGGAAGAAAGAAATGCTGATTAATTTCCCTTGGTATATGTTAAACA	978
Qy	1018	TTGGTATCCGAATTCAGAGCCAGAACCAACCAAGTCCACAAACCGACTCCGGAACCTAGTCC	1077
Db	979	TTGGTATCCGAATTCAGAGCCAGAACCAACCAAGTCCACCAATCGACTCCGGAACCTAGTCC	1038
Qy	1078	AGGCGCGCAACTTGACCAAAATCTTAAATAGACTCAAAATCTTC-----TTT	1125
Db	1039	AAATCGCGCAACTTGACCAAAATCTTCAACAGCTCCCAAGCAATTCAAATGATGAGAATT	1098
Qy	1126	GCTTATGCTAGCTGTACGAAGAAATTTGGGGAAGATATGATTTGCAAGAAAAAGGCATCTC	1185
Db	1099	GGTCAAGAAAGCTGTTGAAAAAGTAAAGCGATGTTATGCTTTTGAGAGAAATGAGATTTCC	1158

```

OY 1186 TCGTATGCTCTTGGGAAAGATTTCACATCTGAACTGTTAAATCTTGAAGCAAGT 1245
    |||||
DB 1159 TCGTATATCCGACGCAAGGATCTTTCAGCAGAAACAGACAGCATTGATAGCAACT 1218
OY 1246 ATCAAAACAGAGAGTGTTCACACACTTAACTGCTAAAAAAGAAATGTTGCTCTCG 1305
DB 1219 GGCACCAAGCAGAAAGTTTATCTCATAGCTAGAGCTAAGAAACTGACTCCCATCTAG 1278
OY 1306 TGACCAAGATTTTATGATTAAGCATATATCTGTTAAGTCTGAGGCTCATTAAGCCTTGT 1365
DB 1279 TGATGAGATTTTCAATTAAGCTTATGACTTACTAGCAGAAATTCACCAAGATTTACT 1338
OY 1366 TGNAAATTAAGGTCGTTAACTTCTGATTTCCAAAGCTTGAACAATTAAGACGCTTGA 1425
DB 1339 TGATATTAAGGTCACAAAGTATTTTGAAGCTTGAATACCTGTTGAAGCACTCA 1398
OY 1426 TGATGAATCGACTAATAAGAAAAATGGTAGATGATTTATGGCAATTCCTAGACCAAT 1485
DB 1399 GGATGTCNCAAGTATTAAGTCAAGTTAGTGAGATATTTCTGCTTCTTAGCTCCGAT 1458
OY 1486 TACCATCCAGAGCACTTGGCAACCAATTCCTCAATTGATGATCTGAAGCAAGT 1545
DB 1459 TCGTATCCAGAACGTTTAGGAAAAACCAATGCGAAATTAACCTACACTGATGATGAGT 1518
OY 1546 TCGTATGCTCAATTAGCTGATGATATCAACGTCAGATGCTTCAATTTTATGATGAACA 1605
DB 1519 TCAAGTAGCCAAAGTTGGCAGGCAAGTACCAACAGAAAGCGTTATATCTTTGATCTCG 1578
OY 1606 TGATATTAACAGTATGAGGAGGATGCTATGTAATCTCAATATGACCATAGCCACTG 1665
DB 1579 TGATATTAACAGTATGAGGAGGATGCTATGTAATCTCAATATGACCATAGCCACTG 1638
OY 1666 GATTGAAAAAGTATACCTTTCTGATTAAGAAAAAGTTGCACTCAAGCTATATCTAAAGA 1725
DB 1639 GATTAAAAAGTATGTTGCTGAAAGCTGAGAGACGCGAGCCGAGCTTATGCTAAAGA 1698
OY 1726 AAAAGGTAATCTTACCTCATCTCCAGACGACATGTTAAAGCAATTCCAAGCTGAGATAG 1785
DB 1659 GAAAGGTTGACCCCTCTTTCAGACAGACATGAGATTCAGAAATCTGAGGCAAAAG 1758
OY 1786 TGCAGCAGCTATTTTAACTGCTGTAAGAGGAAAAAGCAATTCGACTGTTGCACTTC 1845
DB 1759 AGCAAAAGCTATCTTCAACCGGTGTAAGAGCACTTAAGAGTGCCACTGATCGATGCC 1818
OY 1846 ATATATGCTGAGCATACAGTTGAGGTTAAACCGCTAATTTGATTTTCCATTAAGA 1905
DB 1819 TTACAATCTTCAATATGATGTAAGAGTCAAAAAAGGTTTAACTATCACTCATTAAGA 1878
OY 1906 TCATTACGATAATTAATTTGCTTGGTTGATGATCACATCAAAAGCTCCAAATGG 1965
DB 1879 CCATTACGATAATCAAAATTTGAGTGGTTTGAAGAGCCCTTATGAGGCACTTAAGG 1938
OY 1966 CTATATCTTGAAGATTTGTTGCGACGATTAAGTACTACGTAGAACACCCCTGACGAAG 2025
DB 1939 GTATATCTTGAAGATCTTTTGGGCACTGTCAAGTACTATGTCGAATCCAAAGAAAG 1998
OY 2026 TCCACATCTTAATGATGAGTGGGCAATGCCAGTACGATGTGTTAGGCAAGAAAGCA 2085
DB 1999 TCCGATTCAGATTAATGTTTGTGTAAGCTGACACATGTTTCAAAAGAAACAAATATGG 2058
OY 2086 CAGTAAGATCCAAATTAAGAACTTCAAAAGCGGATGAAGAGCCAGTAAGAAAGCACTGCG 2145
DB 2059 TCAACTGATACCAATCAAAAGGAAACCAAGGAGGAGAAACCTCAGACGAAAAACC 2118
OY 2146 TGAGCCAGAGTCCCTCAAGTAGAGA 2171
DB 2119 TGAGGAAGAAACCCCTCGAAGAGAGA 2144

```

RESULT 22
BD063279
LOCUS BD063279 2290 bp DNA linear PAT 27-AUG-2002

```

DEFINITION Streptococcus pneumoniae antigens and vaccines.
ACCESSION BD063279
VERSION BD063279.1 GI:22608882
KEYWORDS JP 2001505415-A/33.
SOURCE unidentifed
ORANISM unidentifed
REFERENCE 1 (bases 1 to 2290)
AUTHORS Kunsch,C.A., Choi,G.H., Johnson,S.L. and Hromocky,J.A.
TITLE Streptococcus pneumoniae antigens and vaccines
JOURNAL Patent: JP 2001505415-A 33 24-APR-2001;
HUMAN GENOME SCIENCES INC
COMMENT
PD 24-APR-2001
PF 30-OCT-1997 JP 1998520667
PR 31-OCT-1996 US 60/029960
PI CHARLES A KUNSCH, GIL H CHOI, SYDOR L JOHNSON, ALEX HROMOCKY PC
C12N15/31, C12N5/18, C12N1/21, C07K14/315, C12Q1/68, A61K39/09, PC
G01N33/569,
PC G01N33/68
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers.
FEATURES
source 1..2290
location/Qualifiers
1..2290
/organism="unidentifed"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
BASE COUNT 766 a 474 c 498 g 547 t 5 others
ORIGIN
Query Match 41.3%; Score 987.6; DB 6; Length 2290;
Best Local Similarity 67.7%; Pred. No. 1.5e-200;
Matches 1481; Conservative 0; Mismatches 645; Indels 60; Gaps 5;
OY 1 TTCTTACGAGTTGGAGCTGATCAAGCTAAGACGTTTAAAGAAAA---TAATCGTGTTC 57
DB 4 TTCTTATGAACCTTGTCGTCCACCAAGCTGTCAGAGTTAAGAAAAAGTCAATCAAGTTTC 63
OY 58 CTATATGATGAAAAACAAGGAGCGCAAAAAACGGAATTTGACTCCTGATGAGGTTAG 117
DB 64 TTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 123
OY 118 CAAGCGTAAGAGATTCATAGTGTGAGCAAAATCGTCATCAAGATAACAGCAAGGCTATGT 177
DB 124 TAAAGGAGGAGGATTCACCGCCGAAACAAATTGTAATCAAGTTTACGATTCAGATTAATGT 183
OY 178 CACTTCAATGCGACCACTATCTATTATTAACATGTTAAGGTTCTTATGACGCTATCAT 237
DB 184 GACCTTCATGAGACCATTAATCACTATTAATGAGCAAGGTTCTTATGATGCCATCAT 243
OY 238 CAGTGAAGATTAATCAATGAAAGATCCAAACTATAAGCTAATAAAGTAGAGTATTTGTTAA 297
DB 244 CAGTGAAGAGCTCTCTATGAAAGATCCGAATTAACAGTTGAAGGATTCAGACATGTCAA 303
OY 298 TGAGGTCAAGGAGGATATGTTATCAAGGTGATGAGTGAAGAAATCACTATGTTTAAAGGA 357
DB 304 TGAATCAAGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 363
OY 358 TGCTGCCACCGGATTAACGTCCTACAAAGAGAAATCAATGACAAAAACAAGACA 417
DB 364 TGCAAGCTCATGCGGATTAATATTTGGACAAAAAGAGATTTAAAGTCGAGAAAGCAAG 423
OY 418 TAGTCACATGTTGAAGGTGAATCTCAAGAAAGATGAGTGTCTGCTTGGCCTTGGACGTTTC 477
DB 424 CAGTCATATATCAT-----AACTCAAGGAGATTAATGATGATGATGATGATGATGATGAT 474
OY 478 GCAAGAGCGTATATCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 537
DB 475 CCAAGAGCGTTTATCAACGAGTATGAGTATATCTTCAATGATATCTGATATCATTTAGAGA 534
OY 538 TACTGTGATGCTTATATGTTCTTCTCAAGATCATTAACATTAATTCCTAAGAAATGA 597

```

```

Db      535  CACGGGTGATGCTTATATGCTTCCTCAGCGGCACTTATCAATTACCTTCTTAAGAATGA 594
Qy      538  GTTATCAGCTAGCAGAGTTGGCTGCTCAGAAAGCCTTCTATCTGTGCGAAGAAATCTGTC 657
Db      595  GTTATCAGCTAGCAGAGTTGGCTGCTCAGAAAGCCTTCTATCTGTGCGAAGAAATCTGTC 629
Qy      658  AAATTCAGAACCTTATCGCCGACAAAATAGCGATTAACCTTCAAGAACAACTGGGTACC 717
Db      630  -----ATTGGAATGGAGAGGAGATCTGCTCTTCTTAAGTTCTAGTATATATG 681
Qy      718  TTCTGTAAGCAATCCAGGAATCAAAATCTAACAAGAACACAGAACACACATACAG 777
Db      662  AAATCAGCTCAACCAAGATTGTCAGAACCAACATCTGACTGTCTCACTCCAACTTATCA 741
Qy      778  TCAAGCAAGTCAAAAGTATGACATGATAGTCTCTTGAAGACGCTTACAAACCTGCTT 837
Db      742  TCA-----AAATCAAGGGGAAACATTTCAAGCCTTTTACGTGAATGTATAGCTAAACCTT 798
Qy      838  GAGTCAACGACATGTAGAACTGTATGCGCTTGTCTTTGATCCAGCACAAAATCACAATGC 897
Db      799  ATCAGAACCCCATGTGGAATCTGATGCGCTTATTTTCGACCCAGGCAATCACAAGTGC 858
Qy      898  AACACTAGAGGTGTGTCAGTCCACAGAGGAGATCATTTACACTTATCCCTTACTCTCA 957
Db      859  AACCCGACAGGTGATGCTGTCTCTCATGTGTAACATTAACACTTATCCCTTATGACA 918
Qy      958  AATGCTGAATTGGAAAGCAATGCTGTATTTATCCCTTCTGTTATCGTTCAAGCA 1017
Db      919  AATGCTGAATTGGAAAGCAATGCTGTATTTATCCCTTCTGTTATCGTTCAAGCA 978
Qy      1018  TTGGGTACAGATTCAAGCCGACAAACCAACAGTCCCAACAGCACTCCGGAACCTAGTCC 1077
Db      979  TTGGGTACAGATTCAAGCCGACAAACCAACAGTCCCAACAGCACTCCGGAACCTAGTCC 1038
Qy      1078  AGCCCGCAGACTCGACCAATCTTAAATAGACTCAAAATCTTC-----TTT 1125
Db      1039  AAGTCGCGCAACTGACCAACCAATCTTCACAGCTCCACCAATCTCAATGTATGACGAAT 1098
Qy      1126  GGTGATGCTGCTGATACGAAAGTTGGGAGGATGTATGATCGAAGAAAGGCACTTC 1185
Db      1099  GGTCAAGAGCTGTTCGAAAGTGAAGGCGATGTATGTCTTTGAGGACAAATGAGCTTTC 1158
Qy      1186  TCGTATATGCTTTGCGAAAGATTTAACATCTGAACTGTAAATACTTGAAGCAAGT 1245
Db      1159  TCGTATATATCCAGGCCAGATCTTCAACAGAAAGAGCAGCATTTGATGACAACT 1218
Qy      1246  ATCAAAACAGAGAGTGTTCACACACTTAACTGCTAAAGAAAGAAATGTTGCTCTCG 1305
Db      1219  GGCACAGCGAGAAAGTTTATCTCATAGCTAGAGCTAAGAAACTGACCTCCACTAG 1278
Qy      1306  TGACCAAGAAATTTATGATTAAGCAATATATCTGTTAACTGAGGCTCAATAAGCCTTGT 1365
Db      1279  TGATGAGAGATTTTCAATTAAGCTTATGACTTACAGAGAAATTCACCAAGATTACT 1338
Qy      1366  TGNAAATTAAGGCTGCTGATTTCTGATTTCCAGCCTTACCAATTTAGAACGCTTGA 1425
Db      1339  TGATATATAAGGTCCAGACAGTGTATTTGAGGCTTTGGAATACTGTGGAACGACTCA 1398
Qy      1426  TGATGAATGCAATATAAGAAATTTGTAAGTATTTATTTGCACTTCTTACGACCAAT 1485
Db      1399  GGATGTGTCAGATGATTAAGTCAAGTTAGTGAGNGATATCTTGTGCTTGTAGCTCGAT 1458
Qy      1486  TACCATCCAGAGCACTTGGCAACCAATTCGAAATTTGAGTATACGAAAGCAAGT 1545
Db      1459  TCGTATCCAGAACGTTTGAAGAAACCAATGCGCAATTTACCTACAGTATGATAGAT 1518
Qy      1546  TCGTATGCTCAATTAGCTGATTAAGTATCAACGTCAGATGTTACATTTTGTATGAACA 1605
Db      1519  TCAAGTAGCAGAGTTGCGAGGACAGTACACACAGAAAGCGTTATATCTTTGATCTCG 1578
Qy      1606  TGATATATCAAGTATGAAGAGATGATATGTAAAGCCTCAATGAGGCATATGATCAGT 1665
Db      1579  TGATATTAACCAAGTATGAAGGGGAGTGCCTATGTAACTCCACATATGACCATAGCACAG 1638

```

```

Qy      1666  GATTGAAAAAGATAGCCTTTCTGATTAAGGAAAAAGTTGAGCTCAAGCTTATCTAAAGA 1725
Db      1639  GATTAAAAAGATAGTTGTCTGAAAGCTGAGAGAGCGGACGCCAGCTTATGCTAAAGA 1698
Qy      1726  AAAAGTATCTTACTTCATCTCCAGACGACAGATTTAAAGCAATCCAACTGAGATAG 1785
Db      1699  GAAAGTTTGAACCCCTCTTGCAGAGACCATCAGATTCAGGAAATATCAGAGCAAAAAG 1758
Qy      1786  TGCAGAGCTATTTCAATCGTGTGAAGGGGAAAAACAATTCACCTCTGTTGACTTCC 1845
Db      1759  AGCAGAGCTATCTCAACCCGCTGAAGACACTTAAGAGTGCCACTGATGATGTC 1818
Qy      1846  ATATATGTTGAGCATACAGTGAAGTTTAAACGCTATTTGATTAATCCATAGA 1905
Db      1819  TTACATCTTCAATATACGTGAAAGTCAAAACGCTATTTAATCATACCTCATATATA 1878
Qy      1906  TCATTACCAATATATTAATTTGCTTGTGATGATCAACATCAAAAGCTCCAAATG 1965
Db      1879  CCATTACCAATATCAATTAATTTGAGTGTGACGAAAGCCTTTATGAGGACCTAAAG 1938
Qy      1966  CTATACCTTGAAGATTGTTTGGCAGCATTAAGTACTACGTAGAACACCTGACGAAG 2025
Db      1939  GTATACCTTGGAGATCTTTTGGCAGCTGTCAAGTACTATGTGCAACATCCAAAGCAAG 1998
Qy      2026  TCCACATCTTAATGATGATGAGGAGCAATGCCAGTGAAGATGTTAGGCAAGAAAGCA 2085
Db      1999  TCCGATTCAGATTAATGTTGTTGGTACCTTGACCACTGTCATGTAAGAAACAAATAG 2058
Qy      2086  CAGTGAAGTCCAAATTAAGAACTTCAAAAGCGATAGAGCCAGTAGAGAAACAACCTTC 2145
Db      2059  TCAAGCTGATTAATCAATCAAAAGGAAACCAAGCAGAGGAAACCTCAGACGAAAAAC 2118
Qy      2146  TGAGCCAGAGTCCCTCAAGTGAAGA 2171
Db      2119  TGAGGAGAAACCCCTCGAGGAAGA 2144

RESULT 23
AX343074
LOCUS       AX343074                2639 bp    DNA    linear    PAT 12-JUN-2002
DEFINITION Sequence 5 from Patent WO0198334.
ACCESSION   AX343074
VERSION     AX343074.1  GI:18152272
KEYWORDS    '
SOURCE      unidentified
            unidentified
            unclassified.
REFERENCE   1
  AUTHORS   Hamel,J., Ouellet,C., Charland,N., Martin,D. and Brodeur,B.
  TITLE     Streptococcus antigens
  JOURNAL   Patent: WO 0198334-A 5 27-DEC-2001;
            SHIRE BIOCHEM INC. (CA)
FEATURES   source
            1..2639
            /organism="unidentified"
            /mol_type="genomic DNA"
            /db_xref="taxon:32644"
            /note="BVH-11-2"
BASE COUNT  889 a      518 c      567 g      665 t
ORIGIN
Query Match      41.1%; Score 980.8; DB 6; Length 2639;
Best Local Similarity 67.8%; Pred. No. 4,4e-199;
Matches 1473; Conservative 0; Mismatches 638; Indels 60; Gaps 5;
Qy      1      TTCTTACGAGTGGGACTGTATCAAGCTAGAAAGGTTAAGGAAA---TAATCGTGTTC 57
Db      173      TTCTTATGAACTTGTGCTGACCAAGCTGCTGTAGTTAAGAAAGTCTAATCGAGTTTC 232
Qy      58      CTATATAGATGAAAAACAAGCAGCAACCAAAACGAGAAATTTGACTCTGATGAGTTAG 117
Db      223      TTATATAGATGATGATCAGGCTGTCTCAAAAGGCAAAAATTTGACACCAAGATGAAGTCAG 292

```

QY 118 CAAGGTGAAGATTCATGCTGAGCAATGTCATCAGATTAACAGACCAAGGCTATGT 177
 Db 293 TAAAGAGAGGGGATCAACGCCGAACAATGTGTATCAAGATTAAGGATTCAGAGGTATGT 352
 QY 178 CACTTCACATGCGCAACCACTATCAATTATTAACATGTAAGGTTCTTATGAAGCTATCAT 237
 Db 353 GACCTTCATGAGACCATTAATTAATTAATGGAAGGTTCTTATGATGACATCAT 412
 QY 238 CAGTGAAGATTTACTCATGAAAGATCCAACTATTAAGCTAAAGATGAGATTTGTTAA 297
 Db 413 CAGTGAAGATTTCTCATGAAAGATCCAAATTAATCAATGTAAGGATTCAGACATTTGCA 472
 QY 298 TGAGTCAAGGGTGGATTTGTTATCAAGGTAGTGAAGAAATCATGTTTACCTTAAAG 357
 Db 473 TGAATTCAGAGGTGGCTATGATTAAGGTAGACGGAATTAATGTTTACCTTAAAG 532
 QY 358 TGCTGCCACGCGATTAACGTCCGTACAAAAGAGAAATCAATGACCAAAAACAAGACA 417
 Db 533 TGCGGCCCATGCGGCAATATTCGGAACAAAAGAGATTAAGCGTCAAGACAGAAACA 592
 QY 418 TAGTCAACATGCTGAAGGTGAATCCCAAGAAACGATGCTGTTGCTTGGCAGCTTC 477
 Db 593 CAGTCATATATCAT-----ACTCAAGAGCAGATTAATGCTGTTGCTGAGCGCAGAGC 643
 QY 478 GCAAGGACGCTATCTACAGATGATGTTATATGCTTAACTGCTTATCATAGAGA 537
 Db 644 CCAAGGACGTTATACACGATGATGGATATATCTTCAATGATGATGATATCAATTAAGAGA 703
 QY 538 TACTGTGATGCTTATATGTTCTTCTCATGAGATCATTACATTCTTAAGATGA 597
 Db 704 CACGGGTGATGCTTATATGTTCTTCTCAAGGACCATTAACATTATCTTAAGAAATGA 763
 QY 598 GTTATACGCTAGCGAGTTGGCTGCTGCAAGACCTTCTATCTGTCAGAGAAATCTGTC 657
 Db 764 GTTATACGCTAGCGAGTTGGCTGCTGCAAGACCT----- 798
 QY 658 AAATTCAGAACTATGCGCCGACAAATAGGATTAACCTTCAAGAACAACTGGGTAC 717
 Db 799 -----ATTGAATGGAAAGCAGGGATCTGCTCTTCTTCAAGTTCTTATATATGC 850
 QY 718 TTCTGTAAGCAATCCAGAGACTACAAATTAACAACAAGCAACAAGCAACACTAAG 777
 Db 851 AAATTCAGTTCAACCAATTTGTCAAGAACCAATCTGACTGCTCACTCAACTATCA 910
 QY 778 TCAAGCAAGTCAAGTAATGACATGATGATCTCTTGAACAAGCTTCAAACTGCTTT 837
 Db 911 TCA---AAATCAAGGGGAAACATTTCAAGCTTTTACGTGAATGTATGCTAAACCTT 967
 QY 838 GAGTCAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 897
 Db 968 ATCAAGAGCGCATGTAGATCTGATGATGATGATGATGATGATGATGATGATGATGATG 1027
 QY 898 AACAGTAGAGGTGTGAGTGCACACAGAGATCATTACACTTCACTCTTCACTCA 957
 Db 1028 AACCGCCAGAGGTGTAGTGTGCTCATGATGATGATGATGATGATGATGATGATGATG 1087
 QY 958 AATGTCGATTTGGAAGACGATGCTGCTGTTATTTCCCTTCTTATGCTTCAACCA 1017
 Db 1088 AATGTCGATTTGGAAGACGATGCTGCTGTTATTTCCCTTCTTATGCTTCAACCA 1147
 QY 1018 TTGGGTACCAATTCAGAGCGACAAACAAGTCCAAACCGACCGGAAACCTAGTCC 1077
 Db 1148 TTGGGTACCAATTCAGAGCGACAAACAAGTCCAAACCTAGTCCGGAACCTAGTCC 1207
 QY 1078 AGGCCCGCAACCTGACCAAAATCTTAAATAGACTCAATTTCTT-----TTT 1125
 Db 1208 AAGTCTGCAACTGACCAAAATCTTCAACAGCTCCAAAGCAATCCAATTGATGAGAAT 1267
 QY 1126 GGTATGCTAGCTGTACGAAAAAGTTGGGAAAGATATGTATTCGAGAAAAAGGCACTTC 1185
 Db 1268 GGTCAAAAGACCTGTTGAAAAAGTGGCGATGTTATGCTTTGAGGAGAAATGAGTTTC 1327

QY 1186 TCGTTATGCTTTGCGAAAGATTTTACCATCTGAACCTGTTAAAAATCTTGAAAGCAATT 1245
 Db 1328 TCGTTATATCCAGCCCAAGATCTTTGACGAGAAACAGCAGCGCATTTGATGACAAT 1387
 QY 1246 ATCAAAACAAGAGTGTTCACACACTTTAATCTGTTAAAAAGAAATGTTGCTCTCG 1305
 Db 1388 GGCACAGAGAAAGTTTATCTCAATAGCTAGAGCTMAAACAATCACCCTCCATCTAG 1447
 QY 1306 TGACCAAGAAATTTATGATTAAGCATTAATCTGTTAAGTCAAGGCTGATAAGGCTTGT 1365
 Db 1448 TGATCGAAATTTTACATTAAGGCTTATGACTTATGACAAAGAAATTCACCAAGTTTACT 1507
 QY 1366 TGNAAATTAAGGGTCTGAATTTCTGATTTCAAGCCTTGAACAATTTATTAAGCCTGAA 1425
 Db 1508 TGATTAATAAGGTGACAAAGTTGTTTGAAGTTTGAAGTTTGAAGTTTGAAGTTTGAAGTT 1567
 QY 1426 TGATGAATCGACTAATAAAGAAATTTGATGATGATTTATTTGCAATTTCTAGCAAT 1485
 Db 1568 GATGATCTCAAGTCAATTAAGCTAAGTGAAGTGAATGATTTCTTCTTGAAGTCCGAT 1627
 QY 1486 TACCATCCAGAGGACTTGGCAACCAATTCCTCAATTTGATATCTGAAGCAAGT 1545
 Db 1628 TCGTATCCAGAGGCTTTAGAAAAACCAATTCCTCAATTTGATATCTGAAGTGAAGT 1687
 QY 1546 TCGTATTCCTCAATTAAGCTGATTAAGTATACAGCTCAGATGTTTATCTTTTGAAGACA 1605
 Db 1688 TCAAGTACCAAGTGTGACAGGCAAGTACACACAGAAAGGTTATATCTTTGATCCCTCG 1747
 QY 1606 TGATATAATCAAGTGAAGAGATGATATGTTAACCTCTCATATGAGGCACTGACTG 1665
 Db 1748 TGATATAACCAAGTATGAGGGGATGCTATGTTAACTCAATATGACCAATACCACTG 1807
 QY 1666 GATTGAAAGATGACCTTTCTGTAAGAAAAAGTTGACGCTCAACCTTATTAAGA 1725
 Db 1808 GATTAAAAAGATGATGTTGCTGAAGCTGAGAGCGGACCGCTTATGCTTAAGA 1867
 QY 1726 AAAAGTATCTTACCTCATCTCCAGACGAGATGTTAAAGCAATCCAACTGAGATAG 1785
 Db 1868 GAAAGTTTGAACCTCTCTTGAACAGACCAAGATTCAGAAATCTGAGCAAAAG 1927
 QY 1786 TGCAGACCTAATTACATGCTGTGAAGGGGAAAAAGAAATTCACCTGTTGACTTCC 1845
 Db 1928 AGCAGAGCTATCTACAAACGCGGAAAGCACTAAGAGGTGCACCTGATCTATGCTC 1987
 QY 1846 ATATATGTTGAGATCAGTTGAGTTAAAAAGGTAATTTGATTTCTCTCAATGA 1905
 Db 1988 TTACAAATCTTCAATATATCTGAGAAATCAAAAAGGATTTAATCACTCAATGA 2047
 QY 1906 TCATTACCATTAATTAATTTGCTTGTGATGATCACAATCAAGGCTCCAAATG 1965
 Db 2048 CCATTACCATTAATCAAAATTTGAGTGTGAGAGGCTTTATGAGGACCTTAAGGG 2107
 QY 1966 CTATACCTTGAAGATTTGTTGCGACGATTAAGTACTAGTGAACAACCTGCGAAG 2025
 Db 2108 GTATAGCTTGAAGATCTTTTGGCGACTGTCATGATCTCAATCCAAAGCAAG 2167
 QY 2026 TCCACATTTCAATGATGAGGCGCAATGCGAGGACATGCTGTAGGCAAGAAAGCA 2085
 Db 2168 TCCGCAATTCAGATATATGTTTGGTAAAGCTAGTGAACAATGTTGTAATAATTAAGGCA 2227
 QY 2086 CAGTGAAGATCCAAATTAAGAACTTCAAGCGGATGAAGAGCCAGTAGAGAAACACTGTC 2145
 Db 2228 CCAAGATAGTAACTCTGATGAGATTAAGAACTGATGAAGTGAAGTGAAGTGAAGTGAAGT 2287
 QY 2146 TGAGCCAGAG 2156
 Db 2288 TGAATCTGATG 2298

RESULT 24
 AE008464 12372 bp DNA linear BCT 13-SEP-2001
 LOCUS
 DEFINITION Streptococcus pneumoniae, R6 section 80 of 184 of the complete

ACCESSION	genome.
VERSION	AE008464
KEYWORDS	AE008464.1 GI:15458511
SOURCE	Streptococcus pneumoniae R6
ORGANISM	Streptococcus pneumoniae R6 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.
REFERENCE	1 (bases 1 to 12372)
AUTHORS	Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S., Dehoff, B.S., Estrem, S., Filtz, L., Fu, D.-J., Fuller, W., Geringer, C., Gilmour, R., Glas, J.S., Knoch, H., Kraft, A., Lagace, R., Leblanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P., McHenry, S., McHenry, M., McLeester, K., Mundy, C., Niclas, T.I., Norris, F.H., O'Garra, M., Peery, R., Robertson, G.T., Rokey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Balz, R.H., Jaskunas, S.R., Rostek, P.R. Jr., Skatrud, P.L. and Glass, J.I.
TITLE	Genome of the bacterium Streptococcus pneumoniae strain R6
JOURNAL	J. Bacteriol. 183 (19), 5709-5717 (2001)
MEETING	21429245
PUBMED	11544234
AUTHORS	2 (bases 1 to 12372)
REFERENCE	Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burgett, S., Dehoff, B.S., Estrem, S., Filtz, L., Fu, D.-J., Fuller, W., Geringer, C., Gilmour, R., Glas, J.S., Hann, A., Khoja, H., Kraft, A., Lagace, R., Leblanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P., McHenry, S., McHenry, M., McLeester, K., Mundy, C., Niclas, T.I., Norris, F.H., O'Garra, M., Peery, R., Robertson, G.T., Rokey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G., Zook, C., Balz, R.H., Jaskunas, S.R., Rostek, P.R. Jr., Skatrud, P.L. and Glass, J.I.
TITLE	Direct Submission
JOURNAL	Submitted (27-JUL-2001) Infectious Diseases Research, Eli Lilly and Company, Lilly Research Labs, Indianapolis, IN 46285-0438, USA
FEATURES	location/Qualifiers
source	1..12372
	/organism="Streptococcus pneumoniae R6"
	/mol_type="genomic DNA"
	/strain="R6"
	/db_xref="taxon:171101"
gene	212..1615
	/gene="yfaA"
	/note="synonym: spr0905"
CDS	212..1615
	/gene="yfaA"
	/codon_start=1
	/transl_table=11
	/label_spr0905
	/product="Cationic amino acid transporter - APC Superfamily"
	/protein_id="AAK9709.1"
	/db_xref="GI:15458512"
	/translation="MESKMT FRTKQVNSLDTEKMRHKLMDLLLGAMVGCVT ITGRAATLAPALVISIVISALCVGSLAFAPASBPATGAYSLVILDEEPA WLAOMTFEMETATISGVASGMAYFGKLQSYGIALPOLANGFNQAGVFLLE LVLAVLSVLINKAKALRNSIIVILKFSALAFVAVGWINIFEDWMSNAPVGLFC IYGSTGIMAGSLMFGFLGFEESISMAVDEKTPKOKIIPRGIVLSLIVILAVLT LVLGVVYSHLVNDVAFAFRSGLSMANVSVAILLTLLTVCISMTAYASRMV SLAADGVPAKFEIKTISKYPRKATITITGLASVVAAGMPLASIAAPCTIATYLI LLAAGLRLRKEKMPRAGEKTPVLPLTILITCLSEMLQYNVTWIAFVALIV GSIIYFYGYGHSTIEE"
gene	1740..2675
	/gene="1mb"
	/note="synonym: spr0906"
CDS	1740..2675
	/gene="1mb"
	/note="putative"
	/codon_start=1
	/transl_table=11
	/label_spr0906
	/product="Hypothetical protein"
	/protein_id="AAK9710.1"

```

gene
/translation="MLBITLSPYGFHILFPOSTIVPHGHYHFIPESDLSAGELAVA
KV"
9577..10125
/gene="phc-truncation"
/note="synonym: spr0910"
9577..10125
/gene="phc-truncation"
/codon_start=1
/transl_table=11
/label=sp0910
/product="Pneumococcal histidine triad protein E
precursor, truncation"
/protein_id="AAK9714.1"
/db_xref="GI:15458517"
/translation="MPVTPTNNGVSAVDGYFNPNDIVRDTGDATVYRHGHYIYIK
SLNNPSPHSNTEVEGSSSVLSNPISLHVHEEDGHPANRISDSGFIYPHG
DHNHYIKVQTGYEALKNKIPSIOSNYPSTFDEKAVLAKVDLADRSIYDKPI
EQROIELALQFTESLKKIYS"
complement(10187..10402)
/gene="spr0911"
complement(10187..10402)
/codon_start=1
/gene="spr0911"
/transl_table=11
/label=sp0911
/product="Hypothetical protein"
/protein_id="AAK9715.1"
/db_xref="GI:15458518"
/translation="MFOLICIAOKTFYFLAICRLLVAIYHVLKQESYNTLQGL
TEIRNPDKTMSVODAIRPAQOQGFNML"
10373..11002
/gene="spr0912"
10373..11002
/gene="spr0912"
/codon_start=1
/transl_table=11
/label=sp0912
/product="Hypothetical protein"
/protein_id="AAK9716.1"
/db_xref="GI:15458519"
/translation="MSKANQLLEQIHIEGIRHSLAEVENTKAAITNKADADLDEVNSQ
VKDYLIRASALQOEKENAEDPOTLVYQKLYDILMHSYLSNNGSDADPFKVA
LLDOLSAKSKDITLLESTKAILVNHQIOSKASASEETSPARNAEANGDTSANEP
NATAESNIEFSDENKESNTRSDPABESTENKTESITTGNGEKVE"
complement(11045..12265)
/gene="pepf"
/note="synonym: spr0913"
complement(11045..12265)
/gene="pepf"
/EC_number="3.4.11.-"

CDS
Query Match 40.3%; Score 962.8; DB 1; Length 1232;
Best Local Similarity 67.4%; Pred. No. 3e-195;
Matches 1461; Conservative 0; Mismatches 648; Indels 60; Gaps 5;

QY 1 TTCTTAGAGTTGGACTGTATCAAGCTAGAACGTTAAGGAAA--TAATCGTTTC 57
Db 2742 TTCTTAGAAGTTGGTGTCTACCAAGCTGTCAAGTTAAGAAAGTCTAATCGATTTC 2801
QY 58 CTATATAGATGAGAAAACAAGCGAAGCAAAACGAGAAATTTGACTCCTGATGAGGTTAG 117
Db 2802 TTATATAGATGTGTATCAGGCTGTCAAAAGGAGCAAAATTTGACACAGATGAAGTACG 2861
QY 118 CAAGCGTAGAGAAATCAATGTGTAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGT 177
Db 2862 TAAAGAGAGAGGATCAACGCCGAACAAATGTTATCAAGATTACGATCAAGGTTATGT 2921
QY 178 CACTTCATGCGCAGCACTATCATTTATACATGATGAAGTTCTTATGACGCTATCAT 237
Db 2922 GACCTTCATGAGACCATTAATCATTAATGAGCAAGGTTCTTATGATGACATCAT 2981

```

```

QY 238 CAGTAGAAGATTTACTCATGAAGAATCCAACTATAGCTTAAAGATGAGATATTGTTAA 297
Db 2982 CAGTAGAAGAACTTCTCATGAAGAATCCGAATATACGTGAAGATCAGACATTGTCAA 3041
QY 298 TGAGGTCAAGGTGTGATATGTTATCAAGGTATGAGAAAATACATGTTTACCTTAAGA 357
Db 3042 TGAATATCAAGGTGTGATGATGATTAAGTAGAGGAAATACATGTTTACCTTAAGA 3101
QY 358 TCGCGCCCAACGCGATACGCTCCGTAACAAAAGAGAAATCAATGACAAACAAAGACA 417
Db 3102 TCGCGCCCAATCGGACAAATATTCGACAAAGAAAGATTAAAGCTGAGAGCAAGAACG 3161
QY 418 TAGTCACATGTGAAGGTGAACTCCAGAAAGATGCTGTCTGTCCTTGACAGCTTC 477
Db 3162 CAGTCATATC-----ACAATCAAGACAGATATATGCTGTGTGTCGACGACAGC 3212
QY 478 GCAAGACGCTATATCTACAGATGATGTTATATCTTTAAATGCTTCTGATATCATAGAGA 537
Db 3213 CCAAGGACGTTATCAACGAGATGAGGTATATCTTCAATGCTTCTGATATCATTTGAAGA 3272
QY 538 TACTGGATGCTTATATTCGTTCTCTCAGAGATCATTAACATTACATTAAGAAATGA 597
Db 3273 TACTGGATGCTTATATTCGTTCTCTCAGAGATCATTAACATTACATTAAGAAATGA 597
QY 598 GTTATCAGCTAGCGAGTTGGCTGTGCGAAGCCTTCTATCTGTGCGAAGAAATCTGTC 657
Db 3333 TTGTCTGCGAGTAATTAGCTGTGCGCAAGCTATTTGAATGG----- 3377
QY 658 AATATCAAGACCTATATGCGGCAAAATAGCGATTAACATTCAAGAAACAACTGGTACC 717
Db 3378 -GAAGCAGGAGATCTCGTCTCTTCAAGTTCTATCTCATTAATGCAAAATCCAGCTCAACCA 3436
QY 718 TTCTTAGCAATCCAGAACTACAAATACATAACAAGCAACAACACATTAACAG 777
Db 3437 GATTGTCAAGAACCAACATCTACTGT-----CACTCCAATTA 3476
QY 778 TCAAGCAAGTCAAGATATGATGATAGTCTCTTGAACAGCTTACAACATGCTCTT 837
Db 3477 TCATCAAAATCAAGGGAACAAATTCAGAGCTTTTACGTAAATGATGTAACCTT 3536
QY 838 GAGTCACGACATATGAAATCTGATGCGCTTGTCTTGAATCCAGACAAATCAACAGTCC 897
Db 3537 ATCAAGAACGCACTATCAAACTGATGCGCTTATTTTGACCCAGCAACGCAACAACTGC 3596
QY 898 AACAGCTAGAGGTTTTCAGTGCAGACGACGAGATCATTAACATCTTCCCTTACTCTCA 957
Db 3597 AACCCCAATGAGTGTCTGTACCGACGAGACATTAATCATCTTATCTTATTCA 3656
QY 958 AATCTGAAATTTGMAAGAAAGAAATCGCTGTATTTATCCCTTGTATCGTTCAACCA 1017
Db 3657 ACTGTCACTTTGMAAGAAATTTGCTGTATTTATCCCTTGTATGTTCAACCA 3716
QY 1018 TTGGGTACCAATTCAGAGCCAGAACCAACCAAGTCCCAACCGACTCCGGAACCTAGTCC 1077
Db 3717 TTGGGTACCAATTCAGAGCCAGAACCAACCAAGTCCCAACCGACTCCGGAACCTAGTCC 3776
QY 1078 AGCGCCGCAACCTCCACCAATCTTAATATAGACTCAAAATCTTC-----TTT 1125
Db 3777 AAGTCCCAACCTCCACCAATCTTCAACGAGCTCCCAAGCAATCAATGATGAGAAAT 3836
QY 1126 GGTATGACGTGTGTAGAAAGTTGGGGAAGATATGTTTCAAGAAAGAGGATCTC 1185
Db 3837 GGTCAAGAAAGCTGTGAAAGATGAGGATGTTATGTTTGAAGAAATGAGATTC 3896
QY 1186 TCGTTATGCTTTTGGAAAGATTTACATCTGAAATCTTAAATCTTGAAGCAAGTT 1245
Db 3897 TCGTTATATCCCAAGCAGATCTTTTCAAGAAAGCAAGCAGAGCAATGATTAACAACT 3956
QY 1246 ATCAAAACAGAGAGTGTTCACACATTTAATCTGTAAGAAAGAAATGTGTCTCTG 1305
Db 3957 GGCACAGCAGAAAGTTTATCTCATAGCTAGAGCTTAAGAAACCTACCTCCATCTAG 4016
QY 1306 TGACCAAGAAATTTATGATTAAGCATATAATCTGTTAATGAGGCTATAAGCCTTGT 1365

```

```

Db      4017 TGATAGAGATTTTCAATAGAGCTTATGACTTACTAGAGAGATTCATCAAGATTACT 4076
Oy      1366 TGNAAATAGAGGCGGTAATTTGATTTTCAGAGCTTAGACAAATTTTGAACGCTTGA 1425
Db      4077 TGATATTAAGTCCAGCAAGTTGATTTGAGCTTGGATACCGTGTGAAAGACTCA 4136
Oy      1426 TGATGATTCGATTAATTAAGAAATTTGATGATGATTTTGGATTCCTAGCAACAT 1485
Db      4137 GGATGCTCAAGTGAATTAAGTCAAGTATGATGATTTCTTGGCTTTTGGCTCCAT 4196
Oy      1486 TACCATCCAGAGCGACTTGGCAACCAATTTCTCAATTTGATGATGATGAGACGAGT 1545
Db      4197 TCGATATCCAGAACGTTTGAAGAAACCAATCGGCAATTTACTACTAGATGATGAGAT 4256
Oy      1546 TCGTATGCTCAATTTAGCTGATTAAGATTAACAAGTACAGTGGTATCATTTTGTAGTAA 1605
Db      4257 TCAAGTAGCAAGTTGGCAGGCAAGTACACAGAAAGCGTTATATCTTTGATCTTG 4316
Oy      1606 TGATATTAATCAAGTATGAAGAGATGATGATTAAGCTTCAATGAGGCTATAGTCACTG 1665
Db      4317 TGATATTAACCAAGTATGAGGAGGATGCTATGATTAATCAATATGATCAATAGCCACTG 4376
Oy      1666 GATTGAAAGATAGGCTTTTCTGATTAAGAAAGAAAGTTGCAAGCTTAACTATTAAGA 1725
Db      4377 GATTAAAAAGATAGTTGCTGTAAGCTGAGAGAGCGGACCGCAAGCTTATGCTTAAGA 4436
Oy      1726 AAAAGTATCTTACCTCCATCTCCAGACGACATGTTAAAGCAATCCAACTGAGATG 1785
Db      4437 GAAAGTTTGAACCCCTCTTCAACAGACATCGAGTTGAGGAAATACGAGGCAAAAG 4496
Oy      1786 TCCAGAGCTTATTTACATCTGTGTGAAGAGGAAAGCAATTCACCTGCTTCACTTCC 1845
Db      4497 AGCAGAGCTATCTTACACCGCTGAAAGACAGTAAAGAGTCCACTGATGCTATGCTGC 4556
Oy      1846 ATATATGTTGACATACAGTTGAGGTTAAAAAGGTTAAATTTGATTTCTTCAATAGA 1905
Db      4557 TTACATCTTCAATATCTGTAAGAGTCAAAAAGGTTATTAATCATCTCATTAATGA 4616
Oy      1906 TCATTAACCAATATTAATTTTGTGTTGATGATGATCAACATCAAAAGCTCCAAATG 1965
Db      4617 CCATTAACCAATATCAAAATTTGAGTGTGTTGACGAAAGGCTTTTGAAGGACCTTAAG 4676
Oy      1966 CTATACCTTGAAGATTTGTTTGCAGATTAAGTACTAGTGAACACCTGAGCAAG 2025
Db      4677 GATATGCTTGAAGATCTTTTGGCAGCTGTCAAGTACTATGTCGAACATCCAAAGCAAG 4736
Oy      2026 TCCACATTTGAATGATGATGGGCAATGCCAGTGAATGATGTTAGGCAAGAAAGCA 2085
Db      4737 TCCGATTCAGATTAATGTTTGGTAAAGCTAGCGACATGTTCAAAAGAAACAAAATG 4796
Oy      2086 CAGTGAAGTCCAAATAGAACTTCAAGCGGATGAAGCGCATGAGAAACACCTTCG 2145
Db      4797 TCAACCTGATCAATCAAAACGAGAAACCAACGAGAGAAACCTCAGACGAAAAACC 4856
Oy      2146 TGAGGCAGA 2154
Db      4857 TGAGGAGA 4865

```

```

RESULT 25
AF340222
LOCUS     AF340222          2166 bp    DNA        linear    SYN 22-MAY-2001
DEFINITION Synthetic construct Phpa-79 (phpa-79) gene, complete cds.
ACCESSION AF340222
VERSION   AF340222.1 GI:13447095
KEYWORDS
SOURCE    synthetic construct
           artificial sequences.
REFERENCE 1 (bases 1 to 2166)
           Zhang, Y., Mast, A.W., Barniak, V., Mountzouros, K., Hostetter, M.K. and
           Green, B.A.
AUTHORS

```

```

TITLE      Recombinant Phpa protein, a unique histidine motif-containing
            protein from Streptococcus pneumoniae, protects mice against
            intranasal pneumococcal challenge
JOURNAL    Infect Immun. 69 (6), 3827-3836 (2001)
MEDLINE    21246685
PUBMED     11349048
REFERENCE  2 (bases 1 to 2166)
AUTHORS    Zhang, Y., Mast, A., Barniak, V., Mountzouros, K., Hostetter, M. and
            Green, B.
TITLE      Direct Submission
JOURNAL    Submitted (25-JUN-2001) Department of Bacteriology, Wveth Lederle
            Vaccines, 211 Bailey Road, West Henrietta, NY 14586, USA
FEATURES   location/Qualifiers
            1..2166
            /organism="synthetic construct"
            /mol_type="genomic DNA"
            /db_xref="taxon:32630"
            /note="derived from Streptococcus pneumoniae Phpa
            sequence"
            1..2166
            /gene="phpa-79"
            1..2166
            /gene="Phpa-79"
            /note="N-terminal truncated Phpa protein; histidine
            motif-containing protein"
            /codon_start=1
            /transl_table=11
            /product="Phpa-79"
            /protein_id="AAK26630.1"
            /db_xref="GI:13447096"
            /translation="MKDPNYOLKSDIVNEIKGVYIKVDKYYVYLKDAAHADNIRT
            KEIRKROKSHNHGSDNOAVVAARAGRYTDDGVIIPASDIIDTGPAYIVPH
            GDHYVITPKNELSSELAARAYNNKGGSSRSSSYNANPAPORLSNNHLYVPT
            YHONGENISILRELYAKPISEKRVESDGLIFDPAQITSTARKAVAPKGVHFIPI
            YEQMSLEKRIARIIPLKRYSNHWVPDSRPOSPSTPESPSQAPAPNPQAPASNP
            IDEKLVSEKRVKGVYFENGVSRYPADLSAETAGIDSKLAKOESLSHKLGAK
            KIDLPSSDRFENKAYDLARIHODLNNKGVDFEALDNLRLKQVPSDKVLVD
            DIALAFIRHBERKXENAOITTYDDEIVAKKLAGKYTEDEGYIFDPDIIISDGA
            VYTPMTHSHWIKKDSLEAERAAOAYAKKGLTPPSDHDODSGNTAKGAFAIYNR
            VKAAKVLDMRPVLYQTYVEYKNSLIIPHVDHYNIKFEKFDGLLEAPAGYTLLED
            LLAIVKYVHEPNERPHSDNGFNASDSVORNRKQAOVNOTERKSEKPEQTERKEE
            TPBERKPESEKPEPTEBESPESEBESPEBESPEBESPEBESPEBESPEBESPEBES
            SNAKETTLGLKNNLLFGTDNNNTIYAEKELALALEKSK"
BASE COUNT 739 a 450 c 462 g 515 t
ORIGIN
Query Match 36.4%; Score 869.4; DB 12; Length 2166;
Best Local Similarity 65.2%; Pred. No. 2.8e-175;
Matches 1436; Conservative 0; Mismatches 627; Indels 141; Gaps 5;
Oy      254 ATGAAAGATCCAACTTAAGCTAAAGATGAGATTTGTTAATGAGGTCAAGGTGCA 313
Db      1 ATGAAAGATCCCAATTAATCAAGTGAAGATTCAGACATTTCAATGAATCAAGGTGCT 60
Oy      314 TATGTTATCAAGGTGATGAGAAATTAATGTTTACCTTAAGAGATGCTCCACGCGGAT 373
Db      61 TATGTTATCAAGGTGATGAGAAATTAATGTTTACCTTAAGAGATGCTCATGCGGAT 120
Oy      374 AACGTCGGTACAAAGAGAAATCAATGACAAAAACAAAGACATAGTCAACGTGTGA 433
Db      121 AATATTCGAGACAAAGAGAGATTAACGTCAAGAGCAGAAACACAGTATATACGGG 180
Oy      434 GGTGGAATCCCAAGAAACGATGCTGCTTGGCTTGGCACGTTGGCAAGAGAGCTATACT 493
Db      181 GGTGATTC-----AACGATCAACAGATGTTGACGACGACGACCAAGAGAGCATATA 234
Oy      494 ACAGATGATGATTAATCTTTATGCTTCTGATATCATAGAGATAGTATGATGCTTAT 553
Db      235 ACGATGATGATTAATCTTTATGCTTCTGATATCATAGATGATGATGATGATGATGAT 294
Oy      554 ATGTTCTCATGAGATCAATTAACATTAATCTTAAGATGATGATTAACGCTACGAG 613
Db      295 ATGTTCTCATGAGGAGACCATTAACATTAATCTTAAGATGATGATTAACGCTACGAG 354

```

QY 614 TTGGCTGTGAGAAAGCTTCCTATCTGTGAGAGAAATCTGTCAATTTCAAGAACTAT 673
 Db 355 TTAGCTGTGAGAAAGCTATTT-----GGAAAT 381
 QY 674 CGCCGACAAAATAGCGATPAACCTTCAAGAACTGGGTAACCTTCTGTAGCAATCCA 733
 Db 382 GGGAAAGCAGGAGATCTGTCTCTTCTTCAAGTTCTAGTTAATGTCAAAATCCAGCTCAACCA 441
 QY 734 GGAATCAAAATATCAACAAGCAACAAGCAACACTTAACGTCAAGCAAGTCAAAAGT 793
 Db 442 AGATTGTAGAGAACCAACATCTGTCTGTCACTCAACCTA---TCATCAAAATTAAGG 498
 QY 794 AATGACATTTAGTCTCTTGAACAGCTCTCAAACTGCTTTGAGTCAAGCAATGTA 853
 Db 499 GAAACATTTCAAGCTTTTACGTGAATGTATGTAAACCTTATCAAGAACGCAATGTG 558
 QY 854 GAATCTGATGGCTTGTCTTGTATCCAGCAACAATCAAGTCAAGCTGAGAGTGT 913
 Db 559 GAATCTGATGGCTTATTTTGAAGCCAGCAATCAAGTCAAGCCGCAAGGTGT 618
 QY 914 GCAGTCCCAACGAGATCATTAACCTTCACTCTCAATGTCTGAATTTGAA 973
 Db 619 GCTGCTCCATGTGTAACCAATTCACCTTATGAACAAAATGTGAATTTGAA 678
 QY 974 GAACGAATCGCTCGTATTTATCCCTTGTATCGTTCAAAACATTTGGGTACCAATCCA 1033
 Db 679 AAAGCAATGCTCGTATTTATCCCTTGTATCGTTCAAAACATTTGGGTACCAATCCA 738
 QY 1034 AGGCGAAGAACCAAGTCCACCAACCGAATCCGGAAGCTAGTCCAGGCCGCAACCTGCA 1093
 Db 739 AGACGAGAACCAACCAAGTCCACCAATCGAATCCGGAAGCTAGTCCAGGCCGCAACCTGCA 798
 QY 1094 CCAATCTTAATAATGACTCA-----ATTCTTCTTGTGTAGTCAAGTGTGTA 1141
 Db 799 CCAATCTTCAACCAAGTCCAGCAACCAATGTAAGAAATGTGTCAAAAGAGCTGT 858
 QY 1142 CGAAAAGTTGGGGAAGATATGTATTCGAGAAAAGGCACTCTCTGTATGTCTTTGCG 1201
 Db 859 CGAAAAGTGGGCAAGTATGTATTCGAGAAAAGGCACTCTCTGTATGTCTTTGCG 918
 QY 1202 AAAGATTTACATCTGAAACCTGTAAATACTTGAAGCAAGTATCAAAAACAAGAGT 1261
 Db 919 AAGATCTTTCAGCGAAGAACAGCAGCAGCATTTGATCAAACTGCGCAACAGAAAGT 978
 QY 1262 GTTTCACACCTTTTACCTGTAAATAAGAAATGTGTCTCTGAGCAAGAAATTTAT 1321
 Db 979 TTATCTCAATAGCTAGAGCTPAAGAAAAGTCACTCCATCTAGTATGAGAAATTTAC 1038
 QY 1322 GATTAAGCATATTAATCTGTAACTGAGGCTCAATAAGCTTGTGTAATAAGGTCGT 1381
 Db 1039 AATAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACTTGATTAATAAGTCA 1098
 QY 1382 AATTCTGATTTTCAAGCTTGAACAATAATTAAGAACTTGAATGATGAAATCACTAAT 1441
 Db 1099 CAAGTTGATTTTGAAGCTTGTGATTAACCTGTGGAACACATCAAGAGATGCCAAAGTAT 1158
 QY 1442 AAAAGAAAATTTGTATGATTTATTTGCAATTTCTTCAAGCAATTAACCATCCAGAGGA 1501
 Db 1159 AAAGTCAAGTTAGTATGATTAATTTCTTCTTGAATCCGATTCGATTCAGAAAGT 1218
 QY 1502 CTTGGCAAAACCAAAATTTCAATTTGATGATTAAGTGAAGCAAGTTCGATTTGCTCAATTA 1561
 Db 1219 TTAGGAAAACCAAAATGCGCAATTTCTAAGCTAGTATGATTAAGTCAAGTCCAAAGTTG 1278
 QY 1562 GCTGATATGATTAACAAGCTAGATGATTAATTTTGTATGAACATGATATATCAAGTAT 1621
 Db 1279 GCAGGCAAGTATCAACAAGAGAGGATTAATTTCTTGTATCCGATGATATTAACAGTAT 1338
 QY 1622 GAAGAGATGATATGATTAAGGCTATATTTGAGGCTATGCTAGTATGATTAAGGATGCT 1681
 Db 1339 GAGGGGATGCTTATGATTAACATATGATGACCATGACCACTGATTAATAAAGATGAT 1398

QY 1682 CTTTCTGATTAAGAAAAAGTTGACGCTCAAGCCTTATATCTAAGAAAAAGTATCTTACT 1741
 Db 1399 TTGTCTGAAGCTGAGAGAGCGGACGACGCTTATCTTAAAGAAAAAGTTTGAACCCCT 1458
 QY 1742 CCATCTCCAGAGCAGATGTTTAAAGCAATCCAACTGAGATATGTCAGCGCTATTTAC 1801
 Db 1459 CTTTGCAGAGCAATTCAGGATTTCAAGAAATCTGAGCAAAAGAGAGAGAGCTATCTAC 1518
 QY 1802 AATCGTGAAGAGGAAAAACGAATTTCCACTCGTTGACTCTCATATATGTTGAGCAT 1861
 Db 1519 AACCGCTGAAGAGAGCTTAAAGAGTGCACCTTATGATGTATGCTTACATTTCAATAT 1578
 QY 1862 ACAGTTAGGTTTAAAAACGCTAATTTGATTAATTTCTTCAAGATCAATTAACATATAT 1921
 Db 1579 ACTGTAGAGTCAAAAACGCTAGTTTATCATCTCATATATGACATTAACATTAACATC 1638
 QY 1922 AATTTGCTGTGTTGATGATCAACATTAACAACTCCAAATGCTTATCTTGGAGAT 1981
 Db 1639 AATTTGAGGTTTGAAGAGGCTTTATGAGGCACTTAAGGGGTATACTTTGAGGAT 1698
 QY 1982 TTGTTGCGAGATTATGATCTAGTAAACACCTGACGAACGTCACATTTCAATAT 2041
 Db 1699 CTTTGGGAGCTGTCAGTACTATGATGACATCAACAGACGTCGCTATTCAGATAT 1758
 QY 2042 GATGGGCAATGCCAGTATGATGTTAGGCAAGAAAGCAACAGTGAAGATCAAT 2101
 Db 1759 GCTTTGTATAGCTATGACGACATGTTCAAGAAACAAAATGCTAAGCTGATACCAAT 1818
 QY 2102 AAGAACTTCAAGCGGATGA----- 2121
 Db 1819 CAACGGAACAAACCAAGCAGAGAAACCTCAGACAGAAAACTGAGAGAAACCCCT 1878
 QY 2122 -----AGACCAAGTAGAG 2134
 Db 1879 CGAGAAAGAAACCGCAAGCAGAAACCAAGTCTCCAAACCAAGAGAACCAAGAA 1938
 QY 2135 GAAACACCTGCTGAGCAGAAAGTCCCTCAAGTATGAGTGAAGAAAGCCCAATC 2194
 Db 1939 GAATCACAGAGAGATCAGAAAGACCTCAGGTCAGACTGAAAGTTGAAGAAACCTG 1998
 QY 2195 AAAGAGCAGAAAGTTTGTCTTGGCAAGTAAAGATCTAGTCTGAAAGCCAAATGCAACA 2254
 Db 1999 AGAGAGCTGAAGTTTACTTGAAGAAATCCAGATCAATTAACAAAGTCCAAAGTCCAAA 2058
 QY 2255 GAACCTAGCTGTTTACGAAATTAATTTGACTCTTCAAAATTAAGTAAACATATATC 2314
 Db 2059 GAGACTTTCACAGATTAATAAATTAATTTTATTTGACACCCGAGACCAACATATAT 2118
 QY 2315 ATGGCAGAGCAGAAAAATTAATTTGCTGCTGTTTAAAGAGTAA 2358
 Db 2119 ATGGCAGAGCTGAAAAATTAATTTGCTTATTAAGAGAGTAA 2162

 RESULT 26
 SPNEU1907/c 75874 bp DNA linear HTG 11-JUL-2001
 LOCUS Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***
 DEFINITION
 ACCESSION AL449929.1 GI:11545154
 VERSION
 KEYWORDS HTG; HTGS_PHASE2.
 SOURCE
 ORGANISM Streptococcus pneumoniae
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 REFERENCE
 1 Dopazo, J., Mendoza, A., Herrero, J., Caldar, F., Humbert, Y.,
 Friedli, L., Guerrier, M., Grand-Schenck, E., Gandin, C., de
 Francesco, M., Poliesi, A., Buell, G., Feger, G., Garcia, E., Peltech, M.,
 and Garcia-Bustos, J.F.
 Annotated draft genomic sequence from a Streptococcus pneumoniae
 type 19F clinical isolate
 JOURNAL Microb. Drug Resist. 7 (2), 99-125 (2001)
 MEDLINE 21335329

PUBMED 11442348
 REFERENCE 2 (bases 1 to 75874)
 AUTHORS Dopezo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A.,
 Humbert, Y., Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de
 Francesco, M., Buell, G., Feger, G., Garcia, E., Pettsch, M., and
 Garcia-Buecos, J. F.
 TITLE Direct Submission
 JOURNAL Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
 Severo Ochoa 2, 28760 Tres Cantos, Spain
 COMMENT * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 FEATURES
 source
 1. 75874
 /organism="Streptococcus pneumoniae",
 /mol_type="genomic DNA",
 /serotype="19F",
 /db_xref="taxon:1313",
 /clone="G54"
 BASE COUNT 22476 a 15318 c 12185 g 25844 t 51 others
 ORIGIN
 Query Match 35.0%; Score 835.8; DB 2; Length 75874;
 Best Local Similarity 67.0%; Pred. No. 3.9e-168;
 Matches 1290; Conservative 0; Mismatches 578; Indels 58; Gaps 5;
 QY 225 ATGAGGCTATCATCTGCTGAGAAATTAATCTGATGAAGATTCGAACATTAAGTAAAGATG 284
 DB 75874 ATGAGGCTATCATCTGCTGAGAAATTAATCTGATGAAGATTCGAACATTAAGTAAAGATG 75815
 QY 285 AGGATATTGTATGAGTCAAGGGTGATGTTATCAAGGATGAGAAATTAAGTAAAGATG 344
 DB 75814 CAGACATTGTCAATGAATCAAGGGTGATGTTATCAAGGATGAGAAATTAAGTAAAGATG 75755
 QY 345 TTTACCTTAAGAGTGTGCGCAAGCGGATTAACGTCGTAACAAAGAGAAATCAATCGAC 404
 DB 75754 TTTACCTTAAGAGTGTGCGCAAGCGGATTAACGTCGTAACAAAGAGAAATCAATCGAC 75695
 QY 405 AAAACCAAGAGATGTCACATCGTGAAGGTGGAATCCAAAGAACATGCTGCTTG 464
 DB 75694 AGAACCAAGAGATGTCACATCGTGAAGGTGGAATCCAAAGAACATGCTGCTTG 75644
 QY 465 CTTGGCAAGTTCGCAAGAGCGCTAATCTAGAGATGATGTTATCTTAAATGCTTGG 524
 DB 75643 CTGCGAGCGAGAGCGCAAGAGCGCTAATCTAGAGATGATGTTATCTTAAATGCTTGG 75584
 QY 525 ATATCATAGAGATGATGCTGATGCTTATATGCTTCATGAGATCAATTAACATTA 584
 DB 75583 ATATCATAGAGATGATGCTGATGCTTATATGCTTCATGAGATCAATTAACATTA 75524
 QY 585 TTCCTTAAGATGATGATCAAGTGAAGTGGCTGCTGCAAGCGCTTCTATCTGCTG 644
 DB 75523 TTCCTTAAGATGATGATCAAGTGAAGTGGCTGCTGCAAGCGCTTCTATCTGCTG 75466
 QY 645 GAGGAATGTCTCAATTCAGAACTTATGCGGACAAATAGCGATTAACATTAAGAA 704
 DB 75465 -----GAGGAGGATTCGCTCTTCTTCAAGTTCTAGTATATGCAAT 75420
 QY 705 CAAACTGGGTACTTCTGTAGCAATCCAGAACTTAACAATTAACAACAACAACA 764
 DB 75419 CCAAGTCAACAAGATGTCTCAAGAAACAACAATTCGATCTG----- 75379
 QY 765 GCAACACTTAAGCTCAAGCAAGTCAAGTGAATGACATGATGCTCTTGAACAAGCTCT 824
 DB 75378 -CACTTCAACTTATCATCAAAATCAAGGGGAAACATTTCAAGCCCTTTTACGTGAATGT 75320
 QY 825 ACAACCTGCTTGAAGTCAAGCAATGTGAATCTGATGCGCTTGTCTTGAATCAAGAC 884
 DB 75319 ATGCTAAACCTTATCAAGAACCGCATGTGAATCTGATGCGCTTATTTTGAACCGAC 75260
 QY 885 AATATCAAGTGAACAGCTAGAGTGTGCAAGTGCACAGAGATCAATTAACATTA 944

DB 75259 AATATCAAGTGAACAGCTAGAGTGTAGCTGTCTCATGTAACCATTAACATTTA 75200
 QY 945 TCCCTTACTTCAAAATGCTGAATGGAAGAAAGAAATGCTGTATTTATCCCTTGGT 1004
 DB 75199 TCCCTTATGAAGAAATGCTGAATGGAAGAAAGAAATGCTGTATTTATCCCTTGGT 75140
 QY 1005 ATCGTTCAAAACCATTTGGGTACAGATTAAGGCGCAAGAACCAAGATCCACAACGACTC 1064
 DB 75139 ATCGTTCAAAACCATTTGGGTACAGATTAAGGCGCAAGAACCAAGATCCACAACGACTC 75080
 QY 1065 CGGAACCTATGTCAGGCGCGCAACCTGCACCAATCTTAAATAGACTCA----- 1115
 DB 75079 CAGAACCTATGTCAGGCGCGCAACCTGCACCAATCTTAAATAGACTCA----- 75020
 QY 1116 ---ATTCTTCTTGGTGTAGCTGATGAGAAAGTGGGAGAGATATGATTCGAAG 1172
 DB 75019 TTGATGAAGAAATGCTCAAGAAAGTGTGCAAGAAAGTGGGAGATGATGCTTTGAGG 74960
 QY 1173 AAAAGGCACTCTCGTTATGCTTTGGCAAAAGATTTACATCTGAACCTGTTAAATTC 1232
 DB 74959 AGAATGAGATTTCTGCTTATATCCAGCCAGAGATCTTACAGCAAAACAGACAGCA 74900
 QY 1233 TTGAAGCAAGTATCAAAACAGAGAGTGTTCACACACTTTAACTGCTAAAGAA 1292
 DB 74899 TTGATGAAGAAATGCTCAAGAAAGTGTGCAAGAAAGTGGGAGATGATGCTTTGAGG 74840
 QY 1293 ATGTTGCTCTCTGTACCAAGATTTTATGATTAAGCATATATCTGTTAATGAGGCTC 1352
 DB 74839 ACTTCCCATGTAGATGAGAAATTTTCAATTAAGCTTATGACTTACGAGAAATTC 74780
 QY 1353 ATTAAGCCTGTTGTTAATAAGAGTGTCTTAATTCGATTTCCAGCCTTAGCAATAT 1412
 DB 74779 ACCAAGATTTACTTATATTAATTAAGTCAACAGTGTATTTGAGGCTTTGATTAACCTGT 74720
 QY 1413 TAGAAGCGTTGAATGATGAATGACTAATTAAGAAATTTGATAGATTAATTTGGCAT 1472
 DB 74719 TGAAGCACTCAAGAGATGCTCAAGTGAATTAAGTCAAGTATGATGATTTCTTGCT 74660
 QY 1473 TCTTGGACCAATTAACCATCAAGGCGACTTGGCAACCAATTTCTCAATTTGATTA 1532
 DB 74659 TCTTGGACCAATTAACCATCAAGGCGACTTGGCAACCAATTTCTCAATTTGATTA 74600
 QY 1533 CTGAAGCAAGATGCTGATGCTCAATTAAGCTGATGATTAATCAAGTCAAGTGTGCA 1592
 DB 74599 CTGATGATGAGATTAAGTGAAGCAAGTGTGCAAGTGTGCAAGTGTGCAAGTGTGCA 74540
 QY 1593 TTTTGTGAATGATGATTAATCAAGTGAAGGAGATGATGATTAAGCTGCTATATG 1652
 DB 74539 TCTTTGATCTCGTGAATTAACAGTGAATGAGGGGATGCTATGATTAATGATTAATG 74480
 QY 1653 GCCATGCTGATGATGATTAAGTGAAGGAGTGTGCTGATTAAGGAAAGTGTGCAAG 1712
 DB 74479 CCCATGCTGATGATTAAGTGAAGGAGTGTGCTGATTAAGGAAAGTGTGCAAG 74420
 QY 1713 CCTATCTAAAGAAAGATTAAGTGAAGGAGTGTGCTGATTAAGGAAAGTGTGCAAG 1772
 DB 74419 CTTATGCTAAAGAAAGATTAAGTGAAGGAGTGTGCTGATTAAGGAAAGTGTGCAAG 74360
 QY 1773 CAATGAGATGATGATGATTAAGTGAAGGAGTGTGCTGATTAAGGAAAGTGTGCAAG 1831
 DB 74359 CTGAGGCAAAAGAGCAAGATTAATTAACAGCGGTGAAGGAGTGAAGGAGTGAAG 74300
 QY 1832 CTCGTTGCACTTCAATATGATGATGATTAAGGAGTGAAGGAGTGAAGGAGTGAAG 1891
 DB 74299 CTTGATGCTATGCTTCAATCTTCAATATCTGATGAAGGAGTGAAGGAGTGAAG 74240
 QY 1892 ATTCTGATTAAGGATTAATCAATTAATTAATTTGCTGATGATGATGATGATGATGAT 1951
 DB 74239 ATACTCTATTAAGCAATTAATCAATTAATTAATTTGCTGATGATGATGATGATGAT 74180
 QY 1952 AAAGCTCCAAATGCTATACCTTGAAGATTTGTTGCAAGATTAATTAATTAATTAAT 2011
 DB 74179 GAGGACCTTAAGGGTATATCTTGAAGATCTTTGGGAGCTGTGAAGTATATGTCGA 74120

QY	2012	CACCGTCGAAACGCCGCAATTCCTTAATGATGATGGGCAGTCCAGTACGATGTGTA	2017
Db	74119	CATCGAAACGAAACGCCGCAATTCAGATATGTTTGTAAACGCTAGGACCAATGTTCA	74066
QY	2072	GGCAAGAAAGACCAACAGTGAAGATCCAAATAAGAACTTCAAGCGGATGAAGCCACTA	21311
Db	74059	AGAAACAAAATATGTCACAGCTGATTCATCAATCAACGAGAAACCTCACAAGAAAACCT	74000
QY	2132	GAGGAA 2137	
Db	73999	GATGAA 73994	
RESULT 27			
SPNEU1929			
LOCUS	SPNEU1929	702 bp	DNA linear HTG 11-JUL-2001
DEFINITION	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.		
ACCESSION	AL449951		
VERSION	AL449951.1	GI:11545176	
KEYWORDS	HTG, HTGS_PHASE2.		
SOURCE	Streptococcus pneumoniae		
ORGANISM	Streptococcus pneumoniae		
	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;		
	Streptococcus.		
REFERENCE	1		
AUTHORS	Dopazo, J., Mendoza, A., Herrero, J., Caldera, F., Pollisi, A., Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de Francesco, M., Pollisi, A., Buell, G., Feger, G., Garcia, E., Peitsch, M., and Garcia-Bustos, J. F.		
TITLE	Annotated draft genomic sequence from a Streptococcus pneumoniae type 19F clinical isolate		
JOURNAL	Microb. Drug Resist. 7 (2), 99-125 (2001)		
MEDLINE	21353329		
PUBMED	11442348		
REFERENCES	2 (bases 1 to 702)		
AUTHORS	Dopazo, J., Mendoza, A., Herrero, J., Caldera, F., Pollisi, A., Humbert, Y., Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de Francesco, M., Buell, G., Feger, G., Garcia, E., Peitsch, M., and Garcia-Bustos, J. F.		
TITLE	Direct Submission		
JOURNAL	Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A., Severe Ochoa 2, 28760 Tres Cantos, SPAIN		
COMMENT	* NOTE: This is a 'working draft' sequence. * This sequence will be replaced * by the finished sequence as soon as it is available and * the accession number will be preserved.		
FEATURES	Location/Qualifiers		
SOURCE	1..702		
	/organism="Streptococcus pneumoniae"		
	/mol_type="genomic DNA"		
	/serotype="19F"		
	/db_xref="taxon:1313"		
	/clone="G54"		
BASE COUNT	233 a 144 c 147 g 176 t	2 others	
ORIGIN			
Query Match	28.2%;	Score 674;	DB 2; Length 702;
Best Local Similarity	97.3%;	Pred. No. 1.7e-133;	
Matches 683;	Conservative	2; Mismatches 17;	Indels 0; Gaps 0;
QY	2119	TTCCTTAAGACGCTATCATCTAGTGAAGATTACTATGAAAGATCCAAACTATTAAGCTAA	278
Db	1	TTCCCTTAAGATGCGATTTTTCAGTGAAGATTACTATGAAGATCCGAACATAAAGCTTAA	60
QY	279	AAGATGAGATATTTCTTAATGAGTGCACAGGCGATATGTTATCAAGGTAGTGGAAAT	338
Db	61	AAGATGAGATATTTCTTAATGAGTGCACAGGCGATATGTTATCAAGGTAGTGGAAAT	120
QY	339	ACTATGTTACTCTTAAGATGCTGCGCCACGCGGATTAAGTCCGTCAAAAGAGAAATCA	398
Db	121	ACTATGTTACTCTTAAGATGCTGCGCCACGCGGATTAAGTCCGTCAAAAGAGAAATCA	180

QY	399	ATCGCAAAAAA	CAAGAGCAT	TGTGCAT	ACATCGTGA	AGGTGGAACT	CCAGAAACA	TG	458
Db	181	ATCGCAAAAAA	CAAGACAT	GTGTCA	ATCATCTGA	AGGTGGAACT	CCAGAAACA	TG	240
QY	459	CTGTTCCTTGG	CAAGCTT	CGCAAGAG	CGCTATACTA	CAGATGAT	GTATATCTT	TATG	518
Db	241	CTGTTCCTTGG	CAAGCTT	CGCAAGAG	CGCTATACTA	CAGATGAT	GTATATCTT	TATG	300
QY	519	CTTCTGATAT	CATATGAG	GATATCTG	TGTATGCTTATAT	CTCTCATG	AGATCATTTA	ACC	578
Db	301	CTTCTGATAT	CATATGAG	GATATCTG	TGTATGCTTATAT	CTCTCATG	AGATCATTTA	ACC	360
QY	579	ATTTCAATTC	CCAAAGAT	GAGTTATC	CAGTACG	AGTGGCTG	CGAAGACCTT	CCAT	638
Db	361	ATTTCAATTC	CCAAAGAT	GAGTTATC	CAGTACG	AGTGGCTG	CGAAGACCTT	CCAT	420
QY	639	CTGGTCGAGA	AAATCTGT	CAAAATTC	CAAGAACTTA	TCGCCGCA	CAAAATAG	CGATTA	698
Db	421	CTGGTCGAGA	AAATCTGT	CAAAATTC	CAAGAACTTA	TCGCCGCA	CAAAATAG	CGATTA	480
QY	699	CAAGAACAA	CTGGGTAC	CTTCTGT	TAAGCAATCC	CAGAACTA	CAATACTA	CAACA	758
Db	481	CAAGAACAA	CTGGGTAC	CTTCTGT	TAAGCAATCC	CAGAACTA	CAATACTA	CAACA	540
QY	759	ACAACAGAA	CACATCA	AGTCAAC	AGTCAAGTAA	AGTAAATG	ACATGATAGT	CTCTGA	818
Db	541	ACAACAGAA	CACATCA	AGTCAAC	AGTCAAGTAA	AGTAAATG	ACATGATAGT	CTCTGA	600
QY	819	AGCTCTCAA	ACTGCTTT	GAGTCA	AGCAGACAT	GTAGATCG	ATGAGCTT	GTGATC	878
Db	601	AGCTCTCAA	ACTGCTTT	GAGTCA	AGCAGACAT	GTAGATCG	ATGAGCTT	GTGATC	660
QY	879	CAGACAAAT	CACAAGT	CGAA	CAGCTAG	AGGTGTTG	CAAGTCC	920	
Db	661	CAGACAAAT	CACAAGT	CGAA	CAGCTAG	AGGTGTTG	CAAGTCC	920	
RESULT 28									
LOCUS	AR219011			2359 bp	DNA		linear	PAT 25-SEP-2002	
DEFINITION	Sequence 243 from patent US 6420135.								
ACCESSION	AR219011								
VERSION	AR219011.1			GI:23319945					
KEYWORDS									
SOURCE	Unknown.								
ORGANISM	Unknown.								
REFERENCE	1 (bases 1 to 2359)								
AUTHORS	Kunesh, C.A., Choi, G.H., Dillon, P.S., Rosen, C.A., Barash, S.C., Fannon, M.R. and Dougherty, B.A.								
TITLE	Streptococcus pneumoniae polynucleotides and sequences								
JOURNAL	Patent: US 6420135-A 243 16-JUL-2002;								
FEATURES	location/Qualifiers								
source	1..2359								
BASE COUNT	786 a	451 c	511 g	611 t					
ORIGIN									
Query Match	27.4%;	Score 653.8;	DB 6;	Length 2359;					
Best Local Similarity	68.3%;	Pred. No. 3.3e-129;							
Matches 1005;	Conservative 0;	Mismatches 413;	Indels 54;	Gaps 5;					
QY	1	TTCTTACAGT	TGGGACTGTAT	CAAGCTAGA	ACGGTTA	AGGAAAA--	TATCGT	GTTC	57
Db	939	TTCTTACAGT	TGGGACTGTAT	CAAGCTAGA	ACGGTTA	AGGAAAA--	TATCGT	GTTC	998
QY	58	CTATATAGAT	GAAAAACA	GGCAGCA	CAAAAA	CGGAGATTTG	ACTCTCTG	TAGGTTAG	117
Db	999	TTATATAGAT	GTGATCAG	CGCTGGTCA	AAAGGCA	AAACCTTGAC	CACACAGATGA	AGTCAG	1058
QY	118	CAAGCGTGA	AGATCAAT	TGCTGAC	CAAAATCTG	TATCAAGTTAC	ACAGCA	CCAGGCTAT	177
Db	1059	TAAAGGAGG	AGGATCA	CGCGCA	AAATTTGTTAT	CAAGATTC	CGATCA	AGGTTAT	1118

```

OY 178 CACTTCACATGGCGACCACTATCATTTATACATGGTAGGTTCCCTTATAGCGTATCAT 237
DB 1119 GACCTCTATGAGACCATTTATCATTTACTATATGGAAGGTTCCCTTATAGATGCCATCAT 1178
OY 238 CAGTGAAGAATTACTCATGTAAGAAATCCAAACTATTAAGCTAAAAAGATGAGGATTTGTTAA 297
DB 1179 CAGTGAAGAAGCTCTCATGTAAGAAATCCGAAATTCATGTTGAAGGATTCAGATTTGTCA 1238
OY 298 TGAGGTCAAGGGTGGATGATGTTATGTAAGGTAGATGGAATACTATGTTTACTTAAAGA 357
DB 1239 TGAATTCAGGGGTTGTTATGTTATGTTATGTTAAAGGTTAACTATGTTTACTTAAAGA 1298
OY 358 TGTCTCCCAAGCGGATTAACGTCCTGTAACAAAAGAGAAATCAATCGACAAAACAGAGA 417
DB 1239 TGCACCTCATGGGATTAATATTCGACAAAAGAGATTAACGTCGACAGACAGAGA 1358
OY 418 TAGTCAACATCTGTGAAGGTGAATTCGAAAGAAACGATGCTGCTTGTGCGACGTTTC 477
DB 1359 CAGTCAATATCACTAACT-----CAAGACAGATTAATGCTGTTGCTGACGACAGAC 1409
OY 478 GCAAGGAGGCTTACTACAGATGATGTTATCTTTATGCTTGTGATATCATAGAGA 537
DB 1410 CCAAGGAGGCTTATTAACGAGATGATGTTATCTTCAATGATGATATCATGAGA 1469
OY 538 TACTGTGATGCTTATGTTCTCTCATGAGATCATTTACATTCATTCCTTAAGATGA 597
DB 1470 CACGGGTATGTTTATGTTCTTCTCAGCGCACATTTACATTCATTCCTTAAGATGA 1529
OY 598 GTTATCAGTACGAGTGGCTGCTGCGAGAACCTTCTTATGCTGCGAAGATCTGTC 657
DB 1530 GTTATCAGTACGAGTGGCTGCTGCGAGAACCTTATGGAATG----- 1574
OY 658 AAATTCAGAACTTATCCCGGACAAATAGCATTAACCTTCAAGAACTAGGATAC 717
DB 1575 -GAGCAGGAGATCTGCTCTTCTTCAAGTTCTGATTAATCAATTCAGCTCAACCA 1633
OY 718 TTCTGTAAAGCAATCCAGAACTAACAATCTAACAAGCAACAAAGCAACTAAG 777
DB 1634 GATTTCTAGAGAACCAACATCTGACTGT-----CACTCCAACTTA 1673
OY 778 TCAAGCAAGTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 837
DB 1674 TCATCAAAATCAAGGGGAAACATTTCAAGCTTTTACGTAATGATGATGATGATGAT 1733
OY 838 GAGTCAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 897
DB 1734 ATCAGAACGCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1793
OY 898 AACAGTGAAGGTTGCGAGTGCACAGGAGATCATTTCACTTATCCCTTATGATGAT 957
DB 1794 AACGCTCAAGGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1853
OY 958 AATGTCGTAATGGAAGAAAGATGCTGCTGATTAATTCCTTCTGTTATGTTTCAACCA 1017
DB 1854 AATGTCGTAATGGAAGAAAGATGCTGCTGATTAATTCCTTCTGTTATGTTTCAACCA 1913
OY 1018 TTGGGTACCAATTCAGGCTCAAGCAACCAAGTCCACACGAGTCCGGAACCTGATG 1077
DB 1914 TTGGGTACCAATTCAGGCTCAAGCAACCAAGTCCACACGAGTCCGGAACCTGATG 1973
OY 1078 AGGCGCGAAGCTGACCAAAATCTTAAATAGCTCAAAATTTCTTCTGTTATGATGAT 1137
DB 1974 AAGTCCGACCAAGCTCCCAAGCAATTCATTAATGAGAA-----ATTGCTCAAGAAC 2027
OY 1138 GGTACGAAAGTTGGGGAAGATATGATTCGAAGAAAAGGCTATCTGTTATGCTTT 1197
DB 2028 TGTTCGAAAGTTAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2087
OY 1198 TGGCAAGATTTACATCTGAAATGTTTAAATCTTGAAGCAAGTTATCAAAACAGA 1257
DB 2088 AGCCAAAGATCTTTAGCAGAAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2147

```

```

OY 1258 GAGTGTTCACACACTTTACTGCTAAAAAGAAATGTTGCTCTGTCGACCAAGATT 1317
DB 2148 AAGTTATCTCTAATAGTAGTAAGAACTAGACTCCATCTAGTATGAGAAATT 2207
OY 1318 TTATGATTAAGCATTAATCTGTTTACTGAGGCTATTAAGCTTGTGTTGAAATTAAGG 1377
DB 2208 TTACATTAAGGCTTATGATCTTACTAGCAAGATTCACAAAGATTTACTGATTAATTAAGG 2267
OY 1378 TCGTAAATCTGATTTCCAGGCTTACAAATTAATTAAGCCTGAAATGATGATGATGAT 1437
DB 2268 TCGACAAAGTGAATTTTGGGCTTTGAGTAACTGTTGAAAGACTCAAGAGATGCTCAAG 2327
OY 1438 TAATTAAGAAAATTTGTAGATGATTTATG 1469
DB 2328 TGATTAAGTCAAGTATGTAAGATATTTCTTG 2359

RESULT 29
BD003923
LOCUS 2359 bp DNA linear PAT 31-JAN-2002
DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.
ACCESSION BD003923.1 GI:18631884
VERSION BD003923.1
KEYWORDS JP 2001501833-A/243.
SOURCE unidentified
ORGANISM unidentified.

REFERENCE 1 (bases 1 to 2359)
AUTHORS Kunsch, C.A., Choi, G.H., Dillon, P.J., Rosen, C.A., Bara, S.C., Fannon, M., and Dougherty, B.A.
TITLE Polynucleotide of Streptococcus pneumoniae and sequence
JOURNAL Patent: JP 2001501833-A 243 13-FEB-2001; HUMAN GENOME SCIENCES INC
COMMENT OS Unidentified
PN JP 2001501833-A/243
PD 13-FEB-2001
PF 30-OCT-1997 JP 1998520718
PR 31-OCT-1996 US 60/029960
PI CHARLES A KUNSCH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, PI STEVEN C BARASH, MICHAEL FANNON, BRIAN A DOUGHERTY, PC C12N15/09, A01K67/027, C07K14/315, C07K16/12, C12N1/15, C12N1/19, PC C12N1/21, PC C12N5/10, C12P21/02, C12Q1/68, G06F17/30, C12N15/00, C12N5/00, PC G06F15/40
CC Strandedness: Double;
CC Topology: linear;
CC Key key
FH key
FT source
FEATURES
source location/Qualifiers
1..2359 /organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
BASE COUNT 786 a 451 c 511 g 611 t
ORIGIN
Query Match 27.4%; Score 653.8; DB 6; Length 2359;
Best Local Similarity 68.3%; Pred. No. 3.3e-129;
Matches 1005; Conservative 0; Mismatches 413; Indels 54; Gaps 5;

```

QY 178 CACTTCACATGGCGACCACTATCATTTATTAACAAGTGTCTCTATGACGCTATCAT 237
 Db 1119 GACCTCTATGAGAGCCATTAATCACTATTAATGCAAGGTTCTTTATGATGCATCAT 1178
 QY 238 CAGTGAAGAATTAATCAATGAAGAATCCAACTATAAGCTAAAGATGAGATATTGTAA 297
 Db 1179 CAGTGAAGAAGCTCTCATGAAGAATCCGAATTAATCAATGAAGAATGAGATATTGTAA 1238
 QY 298 TGAGGTCAAGGTGTGATATGTTATCAAGTATGATGGAATAATCTATGTTTACCTTAAGA 357
 Db 1239 TGAATTAAGAGGTGTATGTTATCAATGAAGAATCCGAATTAATCAATGTTTACCTTAAGA 1298
 QY 358 TGCTGCCACGCGGATAAGTCCGTACAAAAGAGAAATCAATCGACAAAAACAAGACA 417
 Db 1299 TGCAGCTATGCGGATTAATATTCGACAAAAGAAAGATTAAAGCTCGAAGCAGGAAG 1358
 QY 418 TAGTCACATCGTGAAGGTGAATCCAAAGAACGATGCTGTGCTTGGCAGCGTTC 477
 Db 1359 CAGTCATATCATTAATCT-----CAAGAGCAGATATGCTGTGCGACGACAGC 1409
 QY 478 GCAAGAGAGCTTACTACAGATGATGTTTATCTTTAATGCTTGTGATATCATAGAGA 537
 Db 1410 CCAAGGAGGTATACAAAGGATGATGGGTATATCTTCAATGATCTGATATCATTTGAGA 1469
 QY 538 TACTGTGATGCTTATCTGTTCTCTCATGAGATCATTAACATTCATTCCTTAAGATGA 597
 Db 1470 CACGGGTATGCTTATATCTGTTCTCTCATGCGGACCATTAATCATTTCTTAAGATGA 1529
 QY 598 GTTATACGCTTACGAGTGTGCTGCTGCAAGACCTCTCATCTGCTGCAAGAAATCTGTC 657
 Db 1530 GTTATACGCTTACGAGTGTGCTGCTGCAAGACCTCATTTGAATG----- 1574
 QY 658 AAATTCAAGAACCTTATGCGCCACAAAATAGCATTAACCTTCAAGAACAAATGGGTACC 717
 Db 1575 -GAAGCAGAGGATCTGCTCTTCTTCAAGTTCATGTTAATGCAATCAGCTCAACCA 1633
 QY 718 TTCTGTAGCAATCCAGAACTCAAAATATTAACAAGAACCAACCAACATCAACAG 777
 Db 1634 GATGTCTGAGAGACCAATCTGACTGT-----CACTCCAACTTA 1673
 QY 778 TCAAGCAAGTCAAAATGAATGATGATGCTCTTGAACAGCTCTAACAACGCTCTT 837
 Db 1674 TCATCAAAATCAAGGGAACAAATTTCAAGCTTTTACGTGAATGATGCTTAACCTT 1733
 QY 838 GAGTCAAGCATGATGATGATGATGCTGCTGCTTCTTATCGACAAATCAACAGTGC 897
 Db 1734 ATCAGAACGCGATGATGATGATGCTGCTTATTTGACCCAGGCAAAATCACAAGTGC 1793
 QY 898 AACAGTGAAGGTGTGAGTGCACACAGAGATCATTAACCACTTCAATCCCTTACTCTCA 957
 Db 1794 AACGCCAGAGGTGATGCTGCTCTCATGTAACCATTAACCTTATTCCTTATGAACA 1853
 QY 958 AATGTCTGAATTTGGAAGAAGATGCTGATTAATTCCTTCTGTTATCTGTTCAACA 1017
 Db 1854 AATGTCTGAATTTGGAAGAAACGAATGCTGATTAATTCCTTCTGTTATCTGTTCAACA 1913
 QY 1018 TTGGTATCAATTTCAAGGCCAAGAACCAAGTCCACACAGGACTCCGGAACCTAGTCC 1077
 Db 1914 TTGGTATCAATTTCAAGGCCAAGAACCAAGTCCACACAGGACTCCGGAACCTAGTCC 1973
 QY 1078 AGGCCCGCAACCTGACCAAAATCTTAATATGACTCAATTCCTTCTGTTATGAGTCACT 1137
 Db 1974 AAGTCCGCAACAGCTCCAGCAATCAATGATGATGA-----ATTGCTCAAGAGAGC 2027
 QY 1138 GGTACGAAAAGTTGGGGAAGATATGATTCGAAGAAAAGGCAATCTCTGTTATGCTT 1197
 Db 2028 TGTTCGAAAAGTGAAGGATGATGCTTTGAGAGAAATGGAAGTTCTCTGTTATATCC 2087
 QY 1198 TGCAGAAAGTTTACATCTGAAAATCTGTTAAATTTTGAAGCAAGTTATCAAAACAAGA 1257
 Db 2088 AGCCAGAGATCTTTCAGAGAAAAGCAGCAGGATGATGATGCAAACTGCGCAAGCAGGA 2147

QY 1258 GAGTGTTCACACACTTTAATCTGCTAAAGAAAAGTGTGCTCTGTCGACCAAGATT 1317
 Db 2148 AAGTTTATCTCATAGCTAGAGAACTTAAGAAAAGTGTGCTCTGTCGACCAAGATT 2207
 QY 1318 TTATGATTAAGCATATTAATCTGTTAATGAGGCTCATTAAGCCTTTGTTGAATTAAGG 1377
 Db 2208 TTACAAATTAAGCTTATGCTTACTAGCAAGAAATCACCAGATTATCTGATTAATAAGG 2267
 QY 1378 TCGTAATTCGATTTCCAGCCTTAGCAAAATTTATTAAGCCTGATGATGATGAC 1437
 Db 2268 TCGCAAGTGTATTTGAGGCTTTGATTAACCTTTGGAAGCACTCAAGATGCTCAAG 2327
 QY 1438 TAATTAAGAAAATTTGGTATGATTAATTTGG 1469
 Db 2328 TGATTAAGTCAAGTATGATGATTAATTTCTTG 2359
 RESULT 30
 AR120328
 LOCUS AR120328 1342 bp DNA linear PAT 16-MAY-2001
 DEFINITION Sequence 181 from patent US 6159469.
 ACCESSION AR120328
 VERSION AR120328.1 GI:14103904
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 1342)
 AUTHORS Choi, G.H., Kunsch, C.A., Barash, S.C., Dillon, P.J., Dougherty, B.,
 Fannon, M.R., and Rosen, C.A.
 TITLE Streptococcus pneumoniae antigens and vaccines
 . JOURNAL Patent: US 6159469-A 181 12-DEC-2000;
 FEATURES location/Qualifiers
 source 1..1342 /organism="unknown"
 BASE COUNT 451 a 270 c 261 g 360 t
 ORIGIN
 Query Match 16.1%; Score 385.4; DB 6; Length 1342;
 Best Local Similarity 65.3%; Pred. No. 7.7e-72;
 Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;
 QY 37 TAAAGAAATTAATGCTGTTCTTATATGATGAAACCAAGCAGCAAAACGAGAA 96
 Db 28 TAAAGCAATTAATGCTGTTCTTATATGATGAGCAGCAGTCAAGTCAAGAAAGTGA 87
 QY 97 TTGACTCTGATGAGGTAGCAAGCGTGAAGAAATCAATGCTGACAAATGCTATCA 156
 Db 88 CTGACACAGACCGGTAGCCAGAAAGAAATTCAGGCTGAGCAAAATTTGTAATCAA 147
 QY 157 GATTAAGACCAAGGCTATGCTCACTTCAATGCGGACCACTATGATTAATTAAGGTAA 216
 Db 148 AATTAAGATCAAGGCTATGATGATGATGATGATGATGATGATGATGATGATGATG 207
 QY 217 GGTCTTATGAGCCTTATCAATGATGATGATGATGATGATGATGATGATGATGATG 276
 Db 208 AGTTCCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 267
 QY 277 AAAAGATGAGATATTTGATTAAGGTCGAAAGGTGATGATTAATCAAGTATGAGAA 336
 Db 268 TAAAGACGCTGATTTGCTCAATGAAGTCAAGGCTGATTAATCAATCAAGTCAAGTGA 327
 QY 337 ATACTATGTTAATCCTTAAGATGCTGCCAGCGGATTAAGTCTGTTCAAAAGAGAAAT 396
 Db 328 ATATTAATGCTCACTGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 387
 QY 397 CAATGCAAAAACCAAGAGATGATCAATCTGTAAGGTTGAATCTCAAGAAAGAGATG 456
 Db 388 CAATGCTCAAAAACCAAGAGATGATCAATCTGTAAGGTTGAATCTCAAGAAAGAGATG 435
 QY 457 TGTGTTGCTTGGCAGCTTGCAGAGAGCGTATATCAAGATGATGATGATGATGATGAT 516
 Db 436 TAATGTTGCTGATGAGAGGCTCTCAGGAGCATATACAAATGATGATGATGATGATG 495

QY 517 TCGTTCTGATATCATAGAGGATCTGTGATGCTTATATCGTTCTCATGAGATCATTA 576
DB 496 TCCAGCTGATATTTATCGAAGATACGGGTATGCTTATATGTTCCATGAGAGGTACATA 555
QY 577 CCATTACATCTCTAAGATGAGTTATCAGCTAGCGAGTTGGCTGTGCGAAGCCTTCT 636
DB 556 TCACACTACATTTCCCAAAAGGATTATCTGCTAGTGAATTAGACGACGCTAAAGACATCT 615
QY 637 ATCTGTGAGGAATCTGTCAAAATTCAGAACCTATGCCGACAAATAGGATTAACAC 696
DB 616 GGCTGGAAAAATATGCAACCGAGTCAGTTAAGCTATTCTTCAACGCTAGTACAA--- 672
QY 657 TTCAAGAACAACTGGGTACTTCTGTAAAGCAATCCAGAACTACAACTACTAACAACAG 756
DB 673 -----TAAACCCCAATCTGT 687
QY 757 CAACAACAGCAACTACACTGATCAAGCAAGTCAAGTAATGACATTGATAGTCTTGA 816
DB 688 AGCAAAAGGATCACTAGCAGCCAGCAATTAATCTGAAATCTCCAGAGCTTTTGA 747
QY 817 ACAGCTTACAACTGCTTTGAGTCAAGACATGATGATGCTGCTTGA 876
DB 748 GGAATCTATGATTTACCTTAGCGCCCAAGCTTACAGTGAATGAGGCTGCTTGA 807
QY 877 TCCAGCACAATCACAAGTCGAACAGCTAGAGTGTGAGTGCACACAGAGATCATTA 936
DB 808 CCTGCTAAGATTTATCAGTCGTACACCAATGAGATGAGTTCGCGATGGGACCATTA 867
QY 937 CCATTTCACTCCCTTACTCTCAAAATGTGGAATTGGAAGCAATGCTGTAATTAATTC 996
DB 888 CCACCTTATTTCTTACAGCAAGCTTTCTGCTTAGAAGAAAGATGCGAATGCTGTC 927
QY 997 CCTTGT 1003
DB 928 TATCAGT 934

RESULT 31
BD063337 1342 bp DNA linear PAT 27-AUG-2002
LOCUS BD063337
DEFINITION Streptococcus pneumoniae antigens and vaccines.
ACCESSION BD063337
VERSION BD063337.1 GI:22608940
KEYWORDS JP 2001505415-A/91.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 1342)
AUTHORS Kunsch, C.A., Choi, G.H., Johnson, S.L. and Hromockyj, A.
TITLE Streptococcus pneumoniae antigens and vaccines
JOURNAL Patent: JP 2001505415-A 91 24-APR-2001;
HUMAN GENOME SCIENCES INC
PN JP 2001505415-A/91
PD 24-APR-2001
PF 30-OCT-1997 JP 1998520667
PR 31-OCT-1996 US 60/029960
PI CHARLES A KUNSCH, GIL H CHOI, SYDNOR L JOHNSON, ALEX HROMOCKYJ PC
C12N15/31, C12N5/18, C12N1/21, C07K14/315, C12Q1/68, A61K39/09, PC
G01N33/569,
PC G01N33/68
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers.

FEATURES
source 1..1342
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

BASE COUNT 451 a 270 c 261 g 360 t
ORIGIN

Query Match

16.1%; Score 385.4; DB 6; Length 1342;

Best Local Similarity 65.3%; Pred. No. 7.7e-72;
Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

QY 37 TAAAGAAATATATCGTGTCTTCTATATAGATGAAGAAACAAAGCAGCAAAAACGAGAA 96
DB 28 TAAAGCAATATATCTGTCTCTTATGTGATGAGCCAGTCAATCTGAAAAAGTGA 87
QY 97 TTTGACTCTGATGAGTGTAGCAACGTAAGAGATCAATGCTGACAAATGCTATCA 156
DB 88 CTGACACCAAGCCAGGTAGCAGAAAGAAAGAAATTCAGCTGTGCAAAATTTATCA 147
QY 157 GATTAACAGACCAAGGCTATGTCATCTCAATGCGACACATATCATTTATCAATGTA 216
DB 148 AATTACAGATCAGGCTATGTAAGCTCAACGGTACCACTATCTATTAATGCGAA 207
QY 217 GGTTCCTTATGACGCTATCATCAGTAAGAAATTTACTATGAAGATCCAACTATAGCT 276
DB 208 AGTTCTTATGATGCGCTTTTATGAAAGAACTTGTATGAAGATCCAACTATCACT 267
QY 277 AAAAGATGAGATATTTGTAATGAGTCAAGGATGATATGTTATCAAGTATGAGAA 336
DB 268 TAAAGACGCTATATTTGCAATGAAGTCAAGGATGTTATATCATCAAGTATGAGAA 327
QY 337 ATACTATGTTTACCTTAAGATGCTGCCACGCGATTAACGTCCTACAAAAGAGAAAT 396
DB 328 ATATATGCTACCTGAAGATGACGCTCATGCTATATGTTGCACTAAAGATGAAT 387
QY 397 CAATGACAAAAACAAGCATATGCAACATCTGTAAGAGTGAATCCAAAGACGATG 456
DB 388 CAATGCTAAAAACAAGCAATGTCMAAGATATATAGAGTTA-----ACTC 435
QY 457 TCGTGTGCTGGCAGCTGCGCAAGACGATATCTCAAGATGATGTTATATCTTAA 516
DB 436 TATGTTGCTGAGAAAGCTCTCAGGAGCATATACGCAATGATGTTATGCTTTAA 495
QY 517 TCGTTCTGATATCATAGAGATCTGTGATGCTTATATGTTCTCTCATGAGATCATTA 576
DB 496 TCCAGCTGATATTTATCGAAGATACGGGTATAGCTTATATGTTCTCATGAGAGTACATA 555
QY 577 CCATTACATCTCTAAGATGAGTTATCAGCTTAGCAGATTTGGCTGTGCGAAGCCTTCT 636
DB 556 TCACACTACATTTCCCAAAAGGATTATCTGCTAGTGAATTAGACGACGCTAAAGACATCT 615
QY 637 ATCTGTGAGGAATCTGTCAAAATTCAGAACCTATGCGGACAAATAGGATTAACAC 696
DB 616 GGCTGGAAAAATATGCAACCGAGTCAGTTAAGCTATTCTTCAACGCTAGTACAA--- 672
QY 657 TTCAAGAACAACTGGGTACTTCTGTAAAGCAATCCAGAACTACAAATACTAACAACAG 756
DB 673 -----TAAACCCCAATCTGT 687
QY 757 CAACAACAGCAACTACACTGATCAAGCAAGTCAAGTAATGACATTGATAGTCTTGA 816
DB 688 AGCAAAAGGATCACTAGCAGCCAGCAATTAATCTGAAATCTCCAGAGCTTTTGA 747
QY 817 ACAGCTTACAACTGCTTTGAGTCAAGACATGATTAATGATGCTGCTTGA 876
DB 748 GGAATCTATGATTTACCTTAGCGCCCAAGCTTACAGTGAATGATGCTGCTTGA 807
QY 877 TCCAGCACAATCACAAGTCGAACAGCTAGAGTGTGAGTGCACACAGAGATCATTA 936
DB 808 CCTGCTAAGATTTATCAGTCGTACACCAATGAGATGAGTTCGCGATGGGACCATTA 867
QY 937 CCATTTCACTCCCTTACTCTCAAAATGTGGAATTGGAAGCAATGCTGTAATTAATTC 996
DB 888 CCACCTTATTTCTTACAGCAAGCTTTCTGCTTAGAAGAAAGATGCGAATGCTGTC 927
QY 997 CCTTGT 1003
DB 928 TATCAGT 934

RESULT 32

AX568730 3117 bp DNA Linear PAT 29-NOV-2002
 LOCUS AX568780 Sequence 1987 from Patent WO02077021.
 DEFINITION AX568780
 ACCESSION AX568780.1 GI:26002456
 VERSION
 KEYWORDS
 SOURCE Streptococcus pneumoniae
 ORGANISM Streptococcus pneumoniae
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 REFERENCE
 1 Maignani, V., Tetteijn, H. and Fraser, C.
 Streptococcus pneumoniae proteins and nucleic acids
 Patent: WO 02077021-A 1987 03-OCT-2002;
 JOURNAL Chiron Spa (IT) : THE INSTITUTE FOR GENOMIC RESEARCH (US)
 FEATURES
 Source
 1. .3117
 /organism="Streptococcus pneumoniae"
 /mol_type="genomic DNA"
 /db_xref="taxon:1313"
 BASE COUNT 1150 a 526 c 612 g 829 t
 ORIGIN
 Query Match 16.1%; Score 385.4; DB 6; Length 3117;
 Best Local Similarity 65.3%; Pred. No. 7.6e-72;
 Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;
 Oy 37 TAAGAAATAATCTCTGTTTCTTATATAGTGAAGAAACAGAGCAGCAAAACCGAGAA 96
 Db 99 TAAGACAATAATCTCTGTTTCTTATAGTGAAGAAACAGAGCAGCAAAACCGAGAA 158
 Oy 97 TTGACCTCTGATGAGGTGAGCAAGCGTGAAGGAATCATGCTGAGCAATCGTATGAA 156
 Db 159 CTGACACCAAGCCAGGTGAGCAAGGAATCATGCTGAGCAATCGTATGAA 218
 Oy 157 GATAACAGACCAAGGCTATGTCATCTTCATGCGACCACTATCTATTAATGATGATA 216
 Db 219 AATTACAGATCAGGCTATGTAAGTCAACGATGACCACTATCTATTAATGAGAA 278
 Oy 217 GCTTCCTTATGACGCTATGATGAGAAAGATTAATCTGAAAGTCCAACTATAGCT 276
 Db 279 AGTTCCTTATGATGCGCTCTTATGAGAACTCTTGAAGATCCAACTATAGCACT 338
 Oy 277 AAAAGATGAGATATGTTAATGAGTCAAGGCGTGAATGTTATCAAGATGATGATA 336
 Db 339 TAAAGACCTGATATGTTCAATGAAGTCAAGGCGTGAATGTTATCAAGATGATGATA 398
 Oy 337 ATACTATGTTACCTTAAAGATGCTGCCACGCGATTAAGTCCGTACAAAAGAGAAAT 396
 Db 399 AATATATGTTACCTGAAAGATGAGCTCATGCTGATATGTTGCACTTAAAGATGAAT 458
 Oy 397 CAATGACAAAACAGAGCATGATCAACATGCTGAAGTGAAGTCCAAAGACGATG 456
 Db 459 CAATGCTCAAAAACAGAGCATGCTCAAGATTAAGAGGTTA-----ACTC 506
 Oy 457 TGCCTTGCCTTGGACAGCTGCGAAGAGCGCATATCAAGTGAATGTTATCTTAA 516
 Db 507 TAAATGCTGTAAGAGCTCTCAGGAGCATATGCAAAATGATGTTATGTTTAA 566
 Oy 517 TGCCTTGCATATCATAGAGATGATGATGCTTATATGTTCTCTCATGAGATCATTA 576
 Db 567 TCCAGCTGATATATTCGAAGATAGGATATGCTTATATGTTCTCTCATGAGATCATTA 626
 Oy 577 CCATTAATCTCTTAAAGATGAGTATCAAGCTTGAAGTGGCTGCTGCAAGACCTTCT 636
 Db 627 TCACACTACATTCACAAACGATTTATCTGCTAGTGAATTAAGCAGCTTAAAGCATCT 686
 Oy 637 ATTCGTGAGAGAAATCTGTAATTAAGAAACCATGCGACAAATTAAGCATTAAC 696
 Db 687 GCGTGAAGAAATATGCAACGAGTCAAGTATGATCTTCAACAGCTAGAGCAA--- 743
 Oy 697 TTCAAGAAACAACCTGGTACCTTCTGTAAGCAATCAGAAACTAACAATTAACACAA 756

Db 744 -----TAACAGCAATCTGT 758
 Oy 757 CAACAAGCAACACTAAGTCAAGAGTCAAGAAATGATGATATGCTCTTGAA 816
 Db 759 AGCAAAAGATCAACTGACAGCAGCAAAATTAATCTGAAATCTCCAGATCTTTGAA 818
 Oy 817 ACAGCTTCAAACTGCTTGTAGTCAAGATGATGAAATCTGATGAGCTTTTGA 876
 Db 819 GAACTTATGATTCACCTAGCGCCCAAGCTTACAGATGATGATGCGCTGCTTTGA 878
 Oy 877 TCCAGCAAAATCAAGTGAACAGTGAAGTGTTCAGTGCACACGAGATCATTA 936
 Db 879 CCTGCTAAGATTAATGCTGATACCAAAATGAGATTCGATCCGATGCGACATTA 938
 Oy 937 CCATTCATCCCTTACCTCAATGCTGAATGGAAGCAATGCTGCTGATATATTC 996
 Db 939 CCATTCATTCCTTAAGCAAGCTTCTGCTTGAAGAAAGATGCCAGATGTCGC 998
 Oy 997 CTTTGT 1003
 Db 999 TATCAGT 1005
 RESULT 33
 AF318956 3120 bp DNA Linear BCT 11-FEB-2001
 LOCUS AF318956
 DEFINITION Streptococcus pneumoniae pneumococcal histidine triad protein E
 precursor (phle) gene, complete cds.
 ACCESSION AF318956
 VERSION AF318956.1 GI:12744745
 KEYWORDS
 SOURCE Streptococcus pneumoniae
 ORGANISM Streptococcus pneumoniae
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 REFERENCE
 1 (bases 1 to 3120)
 Adamou, J.E., Heinrichs, J.H., Erwin, A.L., Walsh, W., Gayle, T.,
 Dormitzer, M., Degnan, R., Brewah, Y.A., Barren, P., Lathigra, R.,
 Langermann, S., Koenig, S. and Johnson, S.
 Identification and characterization of a novel family of
 pneumococcal proteins that are protective against sepsis
 Infect. Immun. 69 (2), 949-958 (2001)
 JOURNAL PUBMED 21101045
 MEDLINE 11599990
 PubMed 21101045
 TITLE 2 (bases 1 to 3120)
 Adamou, J.E., Heinrichs, J.H., Erwin, A.L., Walsh, W., Dormitzer, M. and
 Johnson, S.
 Direct Submission
 Submitted (03-NOV-2000) Molecular Microbiology, MedImmune, Inc., 35
 West Watkins Mill Road, Gaithersburg, MD 20878, USA
 REFERENCE
 AUTHORS
 FEATURES
 Source
 1. .3120
 /organism="Streptococcus pneumoniae"
 /mol_type="genomic DNA"
 /serotype="4"
 /db_xref="taxon:1313"
 1. .3120
 /gene="phle"
 1. .3120
 /gene="phle"
 /note="phle; protective pneumococcal antigen"
 /codon_start=1
 /transl_table=1
 /product="pneumococcal histidine triad protein E
 precursor"
 /protein_id="AAK06761.1"
 /db_xref="GI:12744746"
 /translation="MKRSKYIAAGSAIVISLCAVALNOHSQENKDNRRVYVDG
 SOSQKSENTPPOVSGKEIQAEQIYIKITDQGYVYSHGDHYVNGKVPYDLFSE
 ELMKDPYQDKADIVNEVKGIIKIVDSKYYVYLKDAHADNVRTRKDEINRQKH
 VKNEKNSVAVARSGRYTNDGVYFNPAIIIEDGNAVIVPHGGHYVIPKSDS
 ASBLAAKAHLIAGKMQPSQISYSTSDNNSTOSVAAGSTSPANKSENIQSLELEY
 DSPSAQRYSDDGVFDPAKISRTPNQVAIPGHQHYHFIYPSKLSALEEKIARMBVI

SGTGSTVSTNAKPNVSSLSGLSSNPSSLTTSKELSSASDGYIINPKDIVEETATAY
 IVRGDHFIYIPKSNQIQPTLPNNSLATPSPLPINPTSHKEHEEDYGDARI I
 ADESGFVMSHDNNHYFPKDLTEBOI KAAQKHLVEKTSNGLDSSHSODAR I
 AKEMKDLCKIEKIAGIMKQYGVKRESIVNKEKNAL IYPHGDHNDPIDHEKPVG
 IGHSHNVELFPREGVAKBENKYYTSGELNVNVLKNSIFNNONFTLANQKRVK
 FSPPELEKULGINMLVKLITPDGKLVESKRVGEVGNLANELDOPVLPQGTFK
 YTIASKDYPEVSYDGFVFTVPTSLAYKMSQTI FYPHADITLARNPOPAVRKGDAL
 VRVDFPHGNAYLENNYKVGELKLP IPKINOGTTRAGNKIIPVTMANAYLDNOSTY I
 VEPILKEKNOTDKPSILPQPKRNKAOEHLKDEKVEBPKTEKEKELSETGSTS
 NSTLEEVPTVDVQEKVAKFAESYGMKLENNVFNMDGTIELYLPGEVKKRMADPTG
 EAPQNGENKPSBENKVSITGTENOPTENKPADSLP EAPANEKPVKPENSTDGMKNPE
 GNVGSDPMLDPALEBPAVDVQEKLEKFTASVIGLSDSVITNMGGTIELRPSGEVI
 KNLSDLIA"

sig_peptide
 1..87
 /gene="phte"
 247..264
 /gene="phte"
 /note="Region: histidine triad"

misc_feature
 442..531
 /gene="phte"
 /note="Region: coiled-coil domain"

misc_feature
 613..630
 /gene="phte"
 /note="Region: histidine triad"

misc_feature
 925..942
 /gene="phte"
 /note="Region: histidine triad"

misc_feature
 1186..1203
 /gene="phte"
 /note="Region: histidine triad"

misc_feature
 1381..1398
 /gene="phte"
 /note="Region: histidine triad"

misc_feature
 1414..1491
 /gene="phte"
 /note="Region: histidine triad"

misc_feature
 1648..1665
 /note="Region: coiled-coil domain"

repeat_region
 522..789
 /rpt_type=direct
 /rpt_type=direct

repeat_region
 1569..1749
 /rpt_type=direct

BASE COUNT 1152 a 526 c 612 g 830 t

ORIGIN

Query Match 16.1%; Score 385.4; DB 1; Length 3120;
 Best Local Similarity 65.3%; Pred. No. 7.6e-72;
 Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

QY 37 TAAAGAAAATATCGTGTTCCTATATAGATGAAACAAAGGACCCAAAAACGAGAA 96
 |||||
 DB 99 TAAAGCAATATACGTCTTATGTGATGACACGCAAGTCAAGCAAAAAAGTAAAA 158
 |||||
 QY 97 TTGACTCCTGATGAGTTAGCAAGCGTGAAGAAATCAATGCTGAGCAATGCTATCA 156
 |||||
 DB 159 CTGACACCAAGCCAGGTAGCCAGAAAGAAATCAAGCTGAGCAATTTGTATCA 218
 |||||
 QY 157 GATAACAGACCAAGGCTATGCACTTCAGATGCGACCACTATCATTTATTAAGTAA 216
 |||||
 DB 219 AATTACAGATCAGGGCTATGTAACGTCAACGCTGACCACTATCATTTATTAAGTAA 278
 |||||
 QY 217 GGTTCCTTGAAGCTATCATCATGAGAAATTAACCAAGAAAGATCCAACTATTAAGT 276
 |||||
 DB 279 AGTTCCTTATGATGCCCTTTAGTAAAGAACTTTGAGAAAGATCCAACTATTAAGT 338
 |||||
 QY 277 AAAAGTAGAGATATTTGTAATGAGTCAAGGCTGATATGTTATCAAGTAGATGAAA 336
 |||||
 DB 339 TAAAGACGCTGATATTTGTAATGAGTCAAGGCTGATATGTTATCAAGTAGATGAAA 398
 |||||
 QY 337 ATATGATGTTTACCTTAAGATGCTGCCACGCGATTAACGTCCTGTAACAAAAGAGAAAT 396
 |||||
 DB 399 ATATGATGTTTACCTTAAGATGCTGCCACGCGATTAACGTCCTGTAACAAAAGAGAAAT 458
 |||||

QY 397 CAATGACAAAAACAAGACATAGTCAACATCGTAGAGTGGAAGCTCCAAAGAACATG 456
 |||||
 DB 459 CAATGCTCAAAAACAAGAACATGTCMAAGATATATAGAAAGTTA-----ACTC 506
 |||||
 QY 457 TCGTGTTCCTTGGACGCTTGCAGAGACGCTATACCTACAGATGATGTTATATCTTAA 516
 |||||
 DB 507 TAAATGCTCTGAGCAAGGCTCAGGAGACGATATACCAAAATGATGTTATGCTTTAA 566
 |||||
 QY 517 TGCTTCATATATCAGAGAGATACCTGATGATGCTTATATCGTTCTCTCATGGAATCATTA 576
 |||||
 DB 567 TCCAGCTATATATTCAGAGAGATACGAGTAAAGCTTATATCGTTCTCTCATGAGGTCACATA 626
 |||||
 QY 577 CCATTAACATTCCTTAAGAAATGAGTATACGACTAGCAGAGTTGGCTGCGCAAGCCCTCCT 636
 |||||
 DB 627 TCACATACATTCCTCAAAAGAGATTTATCTGCTATGTAATTTAGCAGACGCTAAAGCACATTT 686
 |||||
 QY 637 ATCTGCTGAGAGAAATCTGTCAAAATTCAGAAACCTATGCGCCGACAAAATAGCGATPACAC 696
 |||||
 DB 687 GGCTGGAATAAATATGCAACCGAGTCAGTTAGCTATTCCTCACACAGCTAGTACAA---- 743
 |||||
 QY 697 TTCAAGAACAAACCTGGTACCTTCTGTAGACATCCAGAACTACAAATACTATACACAG 756
 |||||
 DB 744 -----TAAACACGCAATCTGT 758
 |||||
 QY 757 CAACAACAGCAACACTAAGACAGTCAAGCAAGTCAAAAGTATGACATTTAGTCTCTGAA 816
 |||||
 DB 759 AGCAAAAGATACATTAAGACAGCCAGCAAAATTAATCTGAAATCTCCAGAGCTTTTGA 818
 |||||
 QY 817 ACAGCTCTCAAACTCGCTTGAAGTCAACGACATGATAGATGAGCTTGTCTTGA 876
 |||||
 DB 819 GGAACCTTATATGATTCACCTAGAGCCCAAGTATGAGTATGAGTGGCTGTCTTGA 878
 |||||
 QY 877 TCCACACAAATCACAAGTCCAGACGCTAGAGGTTGACAGTCCACACGAGATCATTA 936
 |||||
 DB 879 CCCTCTAAGATTATCAGCTGACACCAAGTGAAGTGGATTCGCATGGCGACCATTA 938
 |||||
 QY 937 CCACCTTCAATCCCTTCTCTCAAAATGTCGAATTTGAGAGCAAGATCGCTGATATTC 996
 |||||
 DB 939 CCACCTTATTCCTTACAGACAGCTTCTGCTTGAAGAAAGATTTCCAGATGCTGC 998
 |||||
 QY 997 CCTTCGT 1003
 |||||
 DB 999 TATCAGT 1005
 |||||

RESULT 34
 AX343070 3120 bp DNA linear PAT 12-JAN-2002
 LOCUS AX343070
 DEFINITION Sequence 1 from Patent WO0198334.
 ACCESSION AX343070
 VERSION AX343070.1 GI:18152268
 KEYWORDS
 ORGANISM
 SOURCE
 unclassified.
 unclassified.
 unclassified.

REFERENCE
 1 Hamel,J., Ouellet,C., Charland,N., Martin,D. and Brodeur,B.
 Streptococcus antigens
 Patent: WO 0198334-A 1 27-DEC-2001;
 JOURNAL SHIRE BIOCHEM INC. (CA)
 FEATURES
 source
 1..3120
 Location/Qualifiers
 /organism="unclassified"
 /mol_type="genomic DNA"
 /db_xref="taxon:32644"
 /note="BVH-3"

BASE COUNT 1152 a 525 c 612 g 831 t

ORIGIN

Query Match 16.1%; Score 385.4; DB 6; Length 3120;
 Best Local Similarity 65.3%; Pred. No. 7.6e-72;
 Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

```

Oy 37 TAAGAAATATATGCTGTTTCTTATATAGATGAAAAAAGAGGACGCAAAAAAGGAGAA 96
Db 99 TAAAGCAATATATCTGTCTCTTATGTGATGCGAGCAATGATCGAAAAAGTGAATA 158
Oy 97 TTTGACTCTGATGAGGTTAGCAAGCGTGAAGAAATCAATGCTGACAAATGTCATCAA 156
Db 159 CTTGACACCAAGCCAGGTTAGCCAGAAAGAAATTCAGGCTGAGCAAAATGTATCAA 218
Oy 157 GATAACACCAAGGCTATGTCATCTTCATGCGACCACTATCTTATTTACAAATGATGA 216
Db 219 AATTACAGATCAGGCTATGTAAACGTCAACGCTGACCACTATCTTATTTAAAGGAA 278
Oy 217 GATTCTTATGACGCTATCATCATGTAAGAAATTAATCTGAAAGATCCAAATATATAGCT 276
Db 279 AGTTCTTATGATGCGCTCTTTAGTGAAGAACTTTGATGAAGATCCAAATATATCACT 338
Oy 277 AAAAGATGAGATATTTGTAATGAGGTCAAGGCTGATATGTTATCAAGTATGATGAAA 336
Db 339 TAAAGACGCTGATATGTCATGATGATCAAGGCTGTTATATCATCAAGTGTGATGAAA 398
Oy 337 ATACTATGTTTACCTTAAGATGCTGCCACGCGGATTAACGTCGTAACAAAAGGAAAT 396
Db 399 ATATTATGTTTACCTGAAGATGACGCTCATCTGATATGTTCCAACTAAAGATGAAT 458
Oy 397 CAATGCAAAAAACAAGAGCATAGTCAACATGCTGAGAGTGGAATCCAAAGAAAGATGG 456
Db 459 CAATGCTAAAAACAAGAACATGCTGAAGATATATAGAAAGTTA-----ACTC 506
Oy 457 TGTCTTGTCTTGGCAGCTTGCAGAGAAGCTATATCTACATGATGTTATATCTTTAA 516
Db 507 TAATGTTCTGTAGCAAGGTCTCAAGGACGATATACGAATAATGATGTTATGTTTAA 566
Oy 517 TGTCTTGTATATCATAGAGATCTGTGATGCTTATATCTGTTCTCATGAGATCATTA 576
Db 567 TCCAGCTATATTTATCGAAGATACCGGTAACTGCTTATATGTTCTCATGAGGCTACTA 626
Oy 577 CCATTACATCTCCTAAGATGAGTTATCAAGCTAGCGAGTGGCTGTGAGAAAGCTTCT 636
Db 627 TCATCTACATCTCCAAAGCGATTTATCTGCTAGTGAATTAAGACAGCTAAAGCATCT 686
Oy 637 ATCTGTGAGAGAAATCTGTCAAAATTCAGAACTATGCGCGACAAATAGCGATTAAC 696
Db 687 GCGTGAAGAAAAATATGCAACCGAGTCAGTTAAGCTATTTCTCAACAGCTAGTGA 743
Oy 697 TTCAAGAACAAATGCGTACCTTCTGTAGCAATCCAGAACTACAATTAACAACAAG 756
Db 744 -----TAAACGCAATCTGT 758
Oy 757 CAACAACAGCAACTAAGCAGTCAAGCAAGTCAAGTATGATGATGATCTTGA 816
Db 759 AGCAAAAGATCAACTAGCAAGCCAGCAATTAATCTGAATTTCTCGAGATCTTTTGA 818
Oy 817 ACAGCTCTACAAACTGCTTGAAGTCAAGCATGATAGATCTGATGCTTGTCTTGA 876
Db 819 GGAATCTATGATTTACCTAGCGCCCAAGCTTACAGTGAATCAGATGCTGCTTGA 878
Oy 877 TCCAGCAAAATCACAAGTCAAGCTAGAGGTTTGCAGTCCACACGAGATCATTA 936
Db 879 CCTGCTAAGATTAATCAAGTCAAGCAAGTGAAGTTCGATTCGCAATGGGACATTA 938
Oy 937 CCATCTATCTCCTTAAGTCAAAATGCTGAATTTGGAAGAAAGAAAGATGATATTC 996
Db 939 CCACTTATTTCTTTACAGCAAGCTTCTGCTTTAGAAAGAAAGATTCAGAAATGCTCC 998
Oy 997 CCTTCT 1003
Db 999 TATCAGT 1005

```

RESULT 35
AX343071
LOCUS AX343071 5048 bp DNA linear PAT 12-JAN-2002

```

DEFINITION Sequence 2 from Patent WO0198334.
ACCESSION AX343071
VERSION AX343071.1 GI:18152269
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS Hamel,J., Quellet,C., Charland,N., Martin,D. and Brodeur,B.
TITLE Streptococcus antigens
JOURNAL Patent: WO 0198334-A 2 27-DEC-2001;
SHIRE BIOCHEM INC. (CA)
FEATURES
    source
        1..5048
            /organism="unidentified"
            /mol_type="genomic DNA"
            /db_xref="taxon:32644"
            /note="BVH-3"
BASE COUNT 1709 a 907 c 1104 g 1328 t
ORIGIN

```

Query Match 16.1%; Score 385.4; DB 6; Length 5048;
Best Local Similarity 65.3%; Pred. No. 7.5e-72;
Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

```

Oy 37 TAAGAAATATATGCTGTTTCTTATATAGATGAAAAAAGAGGACGCAAAAAAGGAGAA 96
Db 1875 TAAAGCAATATATCTGTCTCTTATGTGATGCGAGCAATGATCGAAAAAGTGAATA 1934
Oy 97 TTTGACTCTGATGAGGTTAGCAAGCGTGAAGAAATCAATGCTGACAAATGTCATCAA 156
Db 1935 CTTGACACCAAGCCAGGTTAGCCAGAAAGAAATTCAGGCTGAGCAAAATGTATCAA 1994
Oy 157 GATAACACCAAGGCTATGTCATCTTCATGCGACCACTATCTATTTACAAATGATGA 216
Db 1995 AATTACAGATCAGGCTATGTAAGTCAACGCTGACCACTATCTATTAATGGA 2054
Oy 217 GGTCTCTTAAGAGCTATCATCACTGTAAGAAATTCATGAAAGATCCAAATTAAGCT 276
Db 2055 AGTTCTTATGATGCTCCCTCTTTAGTGAAGAACTTTGATGAAGATCCAAATTAAGCT 2114
Oy 277 AAAAGATGAGATATTTGTAATGAGGTCAAGGCTGATATGTTATCAAGTATGAGAAA 336
Db 2115 TAAAGACGCTGATATTTGATGATGAAGTCAAGGCTGTTATATCAAGTGTGATGAAA 2174
Oy 337 ATACTATGTTTACCTTAAGATGCTGCCACGCGGATTAAGCTTCCGTTCAAAAAGGAAAT 396
Db 2175 AATTATGTTTACCTGTAAGATGCTGATGCTGATGATGTTGAACTAAAGATGAAT 2234
Oy 397 CAATGCAAAAAACAAGAGCATAGTCAACATGCGTAAGGTGGAAGTCCAAAGAAAGATGG 456
Db 2235 CAATGCTAAAAACAAGAACATGCTCAAGAAATTAAGAAAGTTA-----ACTC 2282
Oy 457 TGTCTTGTCTTGGCAGCTTGCAGAGAAGCTATATCAAGATGATGTTATCTTTAA 516
Db 2283 TAATGTTGTGTAGCAAGGTCTCAGGACGATATACGCAAAATGATGTTATGCTTTAA 2342
Oy 517 TGTCTTGTATATCATAGAGATCTGTGATGCTTATATGTTTCTCATGAGATCATTA 576
Db 2343 TCCAGCTGATATTTATCAAGATGAGGTTAATGCTTATATGCTTCTCATGAGGCTACTA 2402
Oy 577 CCATTACATCTTAAGATGAGTATCAGCTAGGAGTGGCTGTCAGAAAGCTTCT 636
Db 2403 TCATCTAATTTCCAAAGCGATTTATTTGCTAGTGAATTAAGACAGCTAAAGCAATCT 2462
Oy 637 ATCTGTGAGAGAAATCTGTCAAAATTCAGAACTATGCGCGACAAATAGCGATTAAC 696
Db 2463 GCGTGAAGAAAAATATGCAACCGAGTCAGTTAAGCTATTTCTCAACAGCTAGTGA 2519
Oy 697 TTCAAGAACAAATGCGGTACCTTGTAGCAATCCAGAACTACAATTAACAACAAG 756
Db 2520 -----TAAACGCAATCTGT 2534

```


FT source 1. .6867
 /organisms='Unidentified'.
 FEATURES
 FT Location/Qualifiers
 1. 6867
 /organism="unidentified"
 /mol_type="genomic DNA"
 /db_xref="taxon:32644"
 BASE COUNT 1896 a 1325 c 1212 g 2433 t 1 others
 ORIGIN

Query Match 16.1%; Score 385.4; DB 6; Length 6867;
 Best Local Similarity 65.3%; Pred. No. 7.4e-72;
 Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

37 TAAAGAAATATCGGTTCTTATATAGATGAAACAAAGGAGCAAAACGAGAA 96
 6687 TAAAGCAATATATCGTGTCTTATATGATGAGCGCCGCTAAGTCAGAAAGTGAAAA 6628
 97 TTTGACTCCTGATGAGTTAGCAACGCTGAGAAATCAATGCTGACAAATGCTATCA 156
 6627 CTGACACCGAGCCAGGTTAGCCAGAAAGAAATTCAGGCTGAGCAATGTATCA 6568
 157 GATAACAGACCAAGGCTATGCTTACATGCGACCACTATCATTTATCAATGTGTA 216
 6567 AATTACAGATCAGGGCTATGTACGCTACACGCTGACCACTATCATTTATTAATGGAA 6508
 217 GGTTCCTATGAGCTATCATGAGTAAGAAATTAATCAATGAAAGATCCAACTATTA 276
 6507 AGTTCCTATGATGCGCTCTTTAGAGAACTCTTGAGAAAGATCCAACTATTA 6448
 277 AAAAGATGAGATATTTGTTAATAGATCAAGGCTGATATGTTATCAAGTAGATGAAA 336
 6447 TAAAGACGCTGATATTTGCAATGAAGTCAAGGCTGTTATATCATCAAGTGCAGAAA 6388
 337 ATACTATGTTTACCTTAAAGATGCTGCCACCGGATTAACGCTGACAAAAGGAAAT 396
 6387 ATATTATGCTTACCTGAAAGATGACGCTATGCTATATGTTTCAACCTAAAGATGAAT 6328
 337 CAATGACAAAACAAAGCATGTCACATGCTGAAGTGAAGTCCAGAAAGATGG 456
 6327 CAATGTCAAAACAAAGCATGTCACATGCTGAAGTGAAGTGAAGTGAAGTGAAGTGA 6280
 457 TCGCTGCTGCTGGACGTTCCGACGACGCTATCTACAGATGATGTTATATCTTTAA 516
 6279 TAAATTTCTGATGAGAGGCTCAGGAGAGATATACGAATGATGTTATGCTTTAA 6220
 517 TCGTTCTGATATCATAGAGATATGCTGATGCTTATATGCTTCTCATGAGATCATTA 576
 6219 TCCAGCTATATATTCGAAGATACGGGTATGCTTATATGCTTCTCATGAGGTCACCTA 6160
 577 CCATTACATTCCTAGATGAGTATTCAGCTAGGAGTTGGCTGCTGAGAAAGCTTCT 636
 6159 TCACATCAATTCCTCAAAAGGATTTATCTGCTAGTAAATTAAGCAGACGCTTAAAGCATCT 6100
 617 ATCTGCGAGGAAATCTGTCAAATTTCAAGAACTATGCGCCGACAAAATAGCGATACAC 696
 6099 GGGCTGGAATAATATGCAACGAGTCAGTTAGTATTTCTTCAACAGCTAGAGACA--- 6043
 697 TTCAAGAACAACTGGGTACCTTCTGTAAAGCAATCCAGAACTACAAAATCTAACACAG 756
 6042 -----TAAACGCAATCTGT 6028
 757 CAACAACGCAACACTTAACGTCAGCAAGTCAAGTAAATGATGATAGTCTTTGAA 816
 6027 AGCAAAAGATCACTACAGACGCAAGCAATTAATCTAAATCTCCAGAGCTTTTAA 5968
 817 ACAGCTTCAAACTGCTTTGATGCAAGCATAGATAGTATGATGCTTTGTTGA 876
 5967 GGAATCTATGATTAACCTAGGCCCCAGTTACAGTGAATAGATGCGCTGCTTTGA 5908
 877 TCCAGCAAAATCAGAGTGAAGCTGAAAGCTTGAAGGTTGCAAGTGCACACGAGATCATTA 936
 5907 CCTGCTAAAGATTAATCAGTCGTAACCAATGGAAGTTCGATTCGCAATGGGACATTA 5848

Qy 937 CCATTATCCCTTACTCTCAATGTCTGAATGGAGAACGATCGCTGATATATCC 996
 Db 5847 CCATTATTCCTTACAGCAAGCTTCTGCTTGAAGAAAGATGCGCAATGCTGCC 5788
 Qy 997 CCTGCT 1003
 Db 5787 TATCAGT 5781

RESULT 38
 AR219123
 LOCUS AR219123 973 bp DNA linear PAT 25-SEP-2002
 DEFINITION Sequence 355 from patent US 6420135.
 ACCESSION AR219123
 VERSION AR219123.1 GI:23320057
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 973)
 AUTHORS Kunsch, C.A., Choi, G.H., Dillon, P.S., Rosen, C.A., Barash, S.C.,
 Fannon, M.R. and Dougherty, B.A.
 TITLE Streptococcus pneumoniae polynucleotides and sequences
 JOURNAL Patent: US 6420135-A 355 16-JUL-2002;
 FEATURES
 source location/Qualifiers
 1. 973
 /organism="unknown"
 BASE COUNT 327 a 178 c 211 g 257 t
 ORIGIN

Query Match 16.0%; Score 381.4; DB 6; Length 973;
 Best Local Similarity 77.5%; Pred. No. 5.6e-71;
 Matches 492; Conservative 0; Mismatches 131; Indels 12; Gaps 2;

1 TTCTTACGAGTTGGAGCTGATATCAAGCTAGAACGTTAAGAAAA---TAACTGTTTC 57
 157 TTCTTATGAGCTTGGAGCTTACCAAGCTGTCAGAGTAAAGATGATCAAGTTGC 216
 Qy 58 CTATATGATGAAAAACAGGAGCGCAAAAAACGAGAAATTTGCTCTGATGAGTTAG 117
 Db 217 TTATATGATGTTGATATGAGCTGTCAGAAAGCAAACTGACACGATGAAGTCAG 276
 118 CAACGTAAGAGATTCATGCTGAGCAAAATGCTCATCAAGTAAACAGACCAAGCTATGT 177
 Db 277 TAAAGGAGGAGATCAACCGCCAAATTTGTTATCAAGTTACGATCAAGTTATGT 336
 Qy 178 CACTTCAATGCGACCACTATCATTTATTAATGATGTTAGGTTCTTATGAGCTATCAT 237
 Db 337 GACCTTCATGAGACCATTAATCATTAATTAATGCAAGGTTCTTATGATGCGCATCAT 396
 Qy 238 CAGTGAAGATTAATCATGAAAGATCCAAACTATTAAGTAAAGATAGGATATTTGTTAA 297
 Db 397 CAGTGAAGAGCTCTCATGAAAGATCCGAATTAATCACTTGAAGGATTCAGCATTTGCAA 456
 Qy 298 TGAGTCAAGGCTGATATGTTATCAAGGTAGATGAAATAATCATGTTTACCTTAAAGA 357
 Db 457 TGAATCAAGGCTGATATGTTATCAAGGTAGATGAAAGTAAATCATGTTTACCTTAAAGA 516
 Qy 358 TGTGCCCCAGCGATTAAGCTCGTACAAAAGAGAAATCAATGACAAAAACAAGACA 417
 Db 517 TGCAGCTCATGCGGATATATTTGGACAAAAGAGATTAATTAAGTCAGAAAGCAGAAAG 576
 Qy 418 TAGTCAACATGCTGAAGTGAAGTCCAAAGAAAGATGCTGCTTGTGCTTTGCAAGCTTC 477
 Db 577 CAGTCAATATCAT-----AACTCAAGGACGATTAATGCTGCTGCAAGCAAGC 627
 Qy 478 GCAAGACGCTATCTACAGATGATGTTATCTTTAATGCTTCTGATATCATAGAGA 537
 Db 628 CCAAGACGTTATCAACGAGATGATGCTATATCTTCAATGATCTGATATCATTTAGGA 687
 Qy 538 TACTGTGATGCTTATATGTTCTCTCATGAGATCATTTACATTTCTTAAGAAAGA 597

Db 688 CACGGGTGATGCTTATATCTTCCTCAGCGGACCATTAACATTCATTCTTAAGAATGA 747
Qy 598 GTTATCAGCTAGCGAGTTGGCTGTCAGAAAGCCT 632
Db 748 GTTATCAGCTAGCGAGTTAGCTGCTGCGAAGCCT 782

RESULT 39
BD004035 973 bp DNA linear PAT 31-JAN-2002
LOCUS Polynucleotide of Streptococcus pneumoniae and sequence.
DEFINITION BD004035
ACCESSION BD004035.1 GI:18631996
VERSION JP 2001501833-A/355.
KEYWORDS unclassified
SOURCE unclassified
ORGANISM unclassified

REFERENCE 1 (bases 1 to 973)
AUTHORS Kunesh, C.A., Choi, G.H., Dillon, P.J., Rosen, C.A., Bara, S.C.,
Fannon, M. and Dougherty, B.A.
TITLE Polynucleotide of Streptococcus pneumoniae and sequence
JOURNAL Patent: JP 2001501833-A 355 13-FEB-2001;
HUMAN GENOME SCIENCES INC

COMMENT OS Unidentified
PN JP 2001501833-A/355
PD 13-FEB-2001
PF 30-OCT-1997 JP 1998520718
PI 31-OCT-1996 US 60/029960
PI CHARLES A KUNESH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, PI
STEVEN C BARASH,
PI MICHAEL FANNON, BRIAN A DOUGHERTY
PC C12N15/09, A01K67/027, C07K14/315, C07K16/12, C12N1/15, C12N1/19,
PC C12N1/21,
PC C12N5/10, C12P21/02, C12Q1/68, G06F17/30, C12N15/00, C12N5/00, PC
G06F15/40
CC Strandedness: Double;
CC Topology: Linear;
CC Key Location/Qualifiers
FH source 1..973
FT Location/Qualifiers
1..973
/organism="Unidentified",
/organism="unclassified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

BASE COUNT 327 a 178 c 211 g 257 t
ORIGIN

Query Match 16.0%; Score 381.4; DB 6; Length 973;
Best Local Similarity 77.5%; Pred. No. 5.6e-71;
Matches 492; Conservative 0; Mismatches 131; Indels 12; Gaps 2;

Qy 1 TTCTTACGAGTTGGAGCTGTATCAAGCTAAGCGTTAAAGAAA--TAACTGTTGTTTC 57
Db 157 TTCTTATGAGCTTGGAGCTGTATCAAGCTGTGAGTAAGAAAGTAACTGAAGTTGC 216
Qy 58 CTATATAGTGAAGAAACAGACGCGAAGAAACGAGATTGAGCTCCCTGAGAGTTAG 117
Db 217 TTATATAGTGTGTAGCTGAGCTGTCTAAAGGCGAAGAACTTGACACCGAGTGAAGTCG 276
Qy 118 CAGGCTGAGAGATCAATGTGAGCAATCGTCAATCAAGATTAACAGACCAAGGCTATGT 177
Db 277 TAAAGGAGGAGGAGATCAAGCGGACCAATTTGTTTCAAGATTACGATCAAGGTTATGT 336
Qy 178 CACTTCACTGGCGACCACTATCATTTATTACAATGTAAGTTGTTTCTTATGACGCTATCAT 237
Db 337 GACCTCTCATGAGACCATTTATCATTAATGAGCAAGTTCCCTTATGATGCATCAT 396
Qy 238 CAGTGAAGATTAATCAAGAAAGATCCAAACATTAAGCTAAAGAAAGTGAAGATTTGTTAA 297
Db 397 CAGTGAAGGCTCTCATGAAGAAAGTCCGAATTAATCAAGTGAAGATTCAGACATTTGTCAA 456
Qy 298 TGAGTCAAGGCTGATATATTATCAAGGTAGATGAAGAAATCTATGTTTACCTTAAGGA 357

Db 457 TGAATCAAGGTTGGTTATGTATCATTAAGGTAAACGGTAATATATGTTTACTTAAGGA 516
Qy 358 TGCTCCCAACCGGATTAACGCTCCGACAAAGAGAAATCAATGCACAAAACAGAGGA 417
Db 517 TGCAGCTATCGGATTAATATTTCGACAAAGAGAAATTAACCTCGAAGCAGAAACG 576
Qy 418 TAGTCACATCGTGAAGGTGGAATCCAAAGAAAGATGCTGTGCTTGCGACGTTTC 477
Db 577 CAGTATATATAT-----TACTCAAGACAGATTAATGCTGTGCTGCAGCCAGAC 627
Qy 478 GCAAGACGCTAATACATACAGATGATGTTATATCTTTAATGCTTGTGATATCATAGAGA 537
Db 628 CCAAGACGTTATACAAACGATGATGGGTATATCTTCAATGATCATCATTAATGAAGA 687
Qy 538 TACTGTGATGCTTATATATGTTTCTTCATGAGATCATTAACATTAATCTTAAGATGA 597
Db 688 CACGGGTGATGCTTATATATGTTTCTTCACAGGCGACCATTAACATTAATCTTAAGATGA 747
Qy 598 GTTATCAGCTAGCGAGTTGGCTGCTGCGAAGCCT 632
Db 748 GTTATCAGCTAGCGAGTTAGCTGCTGCGAAGCCT 782

RESULT 40
AX343078 2528 bp DNA linear PAT 12-JAN-2002
LOCUS Sequence 9 from Patent WO0198334.
DEFINITION AX343078
ACCESSION AX343078.1 GI:18152273
VERSION AX343078.1
KEYWORDS unclassified
SOURCE unclassified
ORGANISM unclassified

REFERENCE 1
AUTHORS Hamel, J., Quelet, C., Charland, N., Martin, D. and Brodeur, B.
TITLE Streptococcus antigens
JOURNAL Patent: WO 0198334-A 9 27-DEC-2001;
SHIRE BIOCHEM INC. (CA)

FEATURES
source Location/Qualifiers
1..2528
/organism="unclassified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/note="bvH-3"

BASE COUNT 960 a 398 c 505 g 665 t
ORIGIN

Query Match 14.5%; Score 347; DB 6; Length 2528;
Best Local Similarity 73.0%; Pred. No. 1.2e-63;
Matches 465; Conservative 0; Mismatches 160; Indels 12; Gaps 1;

Qy 37 TAAAGAAATATATGTTGTTTCTTATATATGATGAAACAAACGACGCAAAAAACGAGAA 96
Db 39 TAAAGCAATATATGTTGTTTCTTATATGATGAGGACGACGCAAGTCAAGTCAAGAAAGTGA 98
Qy 97 TTTGACTCTTATGAGTTAGCAACGCGTGAAGAAATCAATGCTGACAAATGCTATCAA 156
Db 99 CTTGACACCAACGAGTTAGCCAGAAAGAAAGAAATTAAGCTGAGCAAAATGTTATCAA 158
Qy 157 GATAACAGACCAAGGCTATGTCACTTACATGGGACACATATCATTTTCAATAGTTAA 216
Db 159 AATTACAGATACGGGCTATGTACGTCACACGGGTGATACATTAATTAATGAGGA 218
Qy 217 GGTTCCTTATGACGCTATCATGAGTGAAGAAATTAATCAAGAAAGTCAAACTATAGCT 276
Db 219 AGTTCCTTATGATGCTCTTATGAGGAACCTGTAATGAAGATCCAACTATCACT 278
Qy 277 AAAAGATAGATATTTGTTAATGAGTCAAGGTTGATGTTATCAAGATGATGAGAA 336
Db 279 TAAACACGCTATATTTGCAATGAAGTCAAGGTTGTTATATCAAGGTTGATGAGAA 338
Qy 337 ATACTATTTAATCTTAAGATGCTGCCACGCGATTAACGTCGTCAAAAAGAGAAAT 396

```

Db      339 ATATTATGTCTACCGAAGATGACGCTCAGTCTATATGTTGCACTAAAGATGAAT 398
Oy      397 CAATGCAAAAACAGAGCATAGTCAATCGTAGAGGTGGAATCCAGAAAGCATGG 456
Db      399 CAATGCTAAAAACAGAAACATGTCAAAGATATGAGAGTTAACTCTAA----- 449
Oy      457 TGTGTTGCTTGGACAGCTTGCAGAGACGCTATCTACATGATGATGTTATCTTTAA 516
Db      450 ---TGTGCTGAGCAAGAGTCTCAGGACGATATCGCAAAATGATGTTATGTTTAA 506
Oy      517 TGTCTTGATATCATAGAGGATCTGTGATGCTTATATCGTTCTCATGAGATCATTA 576
Db      507 TCCAGCTGATATTTATCGAAGATACCGGTAAAGCTTATATCGTTCTCATGAGGCTACTA 566
Oy      577 CCATTACGTTCTTAGAGTAGTGTTCAGTCAGCAGTGGCTGCTGCAAGAGCTTCTCT 636
Db      567 TCACACTACATTCCTCAAAACGATTATTCGTAGTGAATTAAGCAGCAAGCATCT 626
Oy      637 ATCTGTCGAGGAATCTGTCAATTCAAGAACCTAT 673
Db      627 GCTTGAAAAAATATGCAACCGAGTCAGTTAAGCTAT 663

RESULT 41
AE007404      10312 bp      DNA      linear      BCT 31-AUG-2001
LOCUS      Streptococcus pneumoniae TIGR4 section 87 of 194 of the complete
DEFINITION      genome.
ACCESSION      AE007404 AE005672
VERSION      AE007404.1 GI:14972478
KEYWORDS
SOURCE
ORGANISM      Streptococcus pneumoniae TIGR4
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
AUTHORS      1 (bases 1 to 10312)
Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D.,
Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J.,
Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D.,
Umayam,L.A., White,O., Salzberg,S.L., Lewis,M.R., Radune,D.,
Holtzapple,E., Khouri,H., Wolf,A.M., Uterback,T.R., Hansen,C.L.,
McDonald,L.A., Feldblyum,T.V., Angiuoli,S., Dickinson,T.,
Hickney,E.K., Holt,I.E., Loftus,B.J., Yang,F., Smith,H.O.,
Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and
Fraser,C.M.
Complete genome sequence of a virulent isolate of Streptococcus
pneumoniae
JOURNAL      Science 293 (5529), 498-506 (2001)
MEDLINE      11463916
PUBMED
REFERENCE
AUTHORS      2 (bases 1 to 10312)
Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D.,
Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J.,
Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D.,
Umayam,L.A., White,O., Lewis,M.R., Radune,D., Holtzapple,E.,
Khouri,H., Wolf,A.M., Uterback,T.R., Hansen,C.L., McDonald,L.A.,
Feldblyum,T.V., Angiuoli,S., Gesunam,P., Hickey,E.K., Holt,I.E.,
Loftus,B.J., Ujwal,M.L., Yang,F., Smith,H.O., Venter,J.C.,
Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and Fraser,C.M.
Direct Submission
Submitted (29-JUN-2001) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
Location/Qualifiers
source
1. 10312
/organism="Streptococcus pneumoniae TIGR4"
/mol_type="genomic DNA"
/strain="TIGR4"
/db_xref="taxon:170187"
144. 1317
/gene="SP1005"
/note="This region contains a gene with one or more
premature stops or frameshifts, and is not the result of a

```

```

gene
CDS
complement (1426..1629)
/gene="SP1006"
complement (1426..1629)
/note="identified by Glimmer2; putative"
/transl_table=1
/codon_start=1
/product="hypothetical protein"
/protein_id="AAK75123.1"
/db_xref="GI:14972479"
/translacion="MICKAQKTYFFFLAICRRLVAIVHLKQESYNPRLOGLTEIR
NPKTFVODAIRFAQOHGFNML"
1612..2241
/gene="SP1007"
1612..2241
/gene="SP1007"
/note="identified by Glimmer2; putative"
/transl_table=1
/codon_start=1
/product="hypothetical protein"
/protein_id="AAK75123.1"
/db_xref="GI:14972480"
/translacion="MSKANQLLEQIHREGIROSLAEVENLKAATNKADADLDEVNSQ
VKDVLTRIASALOEKENAEODPOTLVLYOKLIDMSLSHYSLENNNGADPDKVA
LLQOLSAKSKDKRLLESTKAILIVNHOLOSASASRETSPARNAEANGTSAENP
MATRESNTERASDENKPSNTRDSKPASTSENETTESSTTGQOEKVE"
complement (2277..3500)
/gene="SP1008"
complement (2277..3500)
/gene="SP1008"
/codon_start=1
/transl_table=1
/product="peptidase C"
/protein_id="AAK75124.1"
/db_xref="GI:14972481"
/translacion="MTYPNLDRFLTYVKVNTSDSHSTTTPSTOSQVDPATNVLPE
MKVGLQNVYVLLNGRAICGLIPANDSELTKIGFISMDPADADNARAEVNVNNG
GVLELGSRGKLDPADKSLERKPGQLLITDGTLLGADKSKIAINRAIVLTLHG
PEIKHCIRFGKFDGDEIGVANKFDEDDVPFAYVVDGSGEGLQETFSAGAEH
HFGQRNHPSTAGQVWVALQALIDFNHQLPNDPRLTGYQGFYHLMVTSVEA
RASVYIRDFPKDAFEARKASMOSIADQNELGSDRYTLNLTDQYNNKEVIEKDMTP
ITAKAGMEDLGIPIPIREGTDSKISFMWIGPPIPIAGCGNNHGREYVSLQTM
ERAVDTIIGVATKYG"
3655..4749
/gene="SP1009"
3655..4749
/gene="SP1009"
/note="identified by match to TIGR protein family HMW
TIGR00109"
/codon_start=1
/transl_table=1
/product="ferrochelatase"
/protein_id="AAK75125.1"
/db_xref="GI:14972482"
/translacion="MKKAILMTFGSPRETIIFGQVADPFTNIRGVPRDHEIOTLYD
NVRIGGTPOKLTROVALVEARLSSEISVFRANKSPFIPVTCOMEDAGDEQCI
CLLEPHYSVSYWGEKPLESKOIFLVIKDWYQEBALLNWMADLAKLIKKEVKQD
SPKVIPEASHVPIFALDFCDPYIDQIFENSKLVAEKGLSSQGYTNMOSESDDIGPW
IKDDVLEYLEQTEHPHYIFVPISEHIEVLFDNDVSCYDLCOEFGVYHRPPMP
NTDSRLIDALVNTVRYVNEQEFKPELPDEETPELVPSDBETKNILAESQDLQWPEFAK
KLIBKGRENVKPYLIKMLKAKKLPKE"
complement (4791..5168)
/gene="SP1010"
complement (4791..5168)
/gene="SP1010"
/note="identified by match to Pfam protein family HMW
PF01741"
/codon_start=1
/transl_table=1
/product="large conductance mechanosensitive channel

```

```

gene      protein MscL
CDS       /protein_id="AAK75126.1"
          /db_xref="GI:14972483"
          /translation="MKMKSLRGNVIDLAVGVVIAAFATVSLVNDITPIL
          NPAKAKVERIAQLSMGVGNFSLAINPIFGTALPFIKIKKAKLTKIEE
          KTDKKPTELEVLQETKALEKK"
          complement(5250..5681)
          /gene="SP1011"
          complement(5250..5681)
          /note="similar to GB:X73124 SP:P39602 PID:413958
          GB:AU009126; identified by sequence similarity; putative"
          /codon_start=1
          /transl_table=11
          /product="Gera family protein"
          /protein_id="AAK75127.1"
          /db_xref="GI:14972484"
          /translation="MKKPIQKPFNNEILAYLFFGVITLVAISRLVIYLSHOLLA
          TALANIGLFAFINDTIVFKOKKNRILRVKESLSTFLDLFLFLVIOFP
          HIGQFVENIKINAETIIAQIFIIINILSKYIPRK"
          /gene="SP1012"
          /db_xref="GI:14972485"
          /translation="MKKLTIRVDADIAVAIYVVLTPPLNAISYGAYQRISEM
          NEMAFYPIYIGVTIGCMIANFESGLDVFVGGSTFLPLSLGWLPFAKYSKQYLF
          NGLIKRDEFFSILFSISMFTIAELINVAELPFLWTGTTGIFPASLIIGAIITIK
          IGRRLDSK"
          /gene="SP1013"
          /db_xref="GI:14972486"
          /translation="MGYVAVVGATGAVGAGOMIKMLESTLPIDKIRYLASARSAGKS
          LKFDODITEETETAFEGVDIALFSAGSSTSAKAPYAVAKAGVVDNNTSPRON
          DPLVPEVNAALAHNGIICPNCSOTOMVALPEPQKGLRITITSTYQASGA
          CMGAILETORELREVLNDGVKPRDLHAELIPSGDGKPIAFNMLPQIDVFTDVT
          YEEMTKETKTKIMEDDSIAVSATCVRIPLVLSAESVYIEKEVAPIEEVAAIAAF
          PGAVLEDVAHQIYPOAINAVGSRDTFVGRIRKDLDAEKGIHMMVSDMLKGAAMS
          VOIAETLHERGIVRPTAELKFLK"
          /gene="SP1014"
          /db_xref="GI:14972487"
          /translation="MSYODLKCKKIITAFITPFHEDSGINPDAIPALIEHLAHTDG
          ILAGTASPLTLDDELELFAVQKVNGRVPILAGVCTNDTSDTEPEVVAEFG
          GRAAGIATPYNKSQSGMYOHFKTIADSLPITIIYIPGRVVEITPEETMLAD
          HENITGVKSCISLANNAYLIERKPEEFLIYTEDGDAPFANMLGADGVISVASHNGD
          ENHEMTAIESDMKAAAIQRKFIKPVNALSYSPPAVKAILNVMGEAPRTPLPL
          VPAPEADAKRIIKVVVDGDEATKATVTVGLRPDY"
          complement(8945..10202)

```

```

/gene="SP1015"
/note="This region contains an authentic frame shift and
is not the result of a sequencing artifact; similar to
GP:2804700; identified by sequence similarity; putative;
IS1167, transposase, authentic frameshift"
BASE COUNT      3258 a      1815 c      2088 g      3151 t
ORIGIN
Query Match      10.6%; Score 253.2; DB 1; Length 10312;
Best Local Similarity 70.5%; Pred. No. 1.3e-43;
Matches 358; Conservative 0; Mismatches 138; Indels 12; Gaps 1;
37  TAAGGAAATATATCGTGTTCCTATATAGAGGAAACAGCGACGCAAAAAACGAGNA 96
   |||
Db    242  TACAGATAAATATCGTGTTCCTATATAGAGGAAACAGCGACGCAAAAAACGAGNA 301
   |||
QY    97  TTTGACCTCGATGAGGTGTTAGCAAGCGTGAAGAAATCATGCTGCAAAATGTCATCA 156
   |||
Db    302  TTTGACCTCGAGCTAGTTAGTCAAAAAGAGAAATCAAGCAAAAATGTTATTVA 361
   |||
QY    157  GATACAGACCAAGGCTATGTCACCTTCATGCGACCACTATCATTTATCAATGTGA 216
   |||
Db    362  GATTACAGATCAAGGTATGTAACCTTCATGCGACATCATTAATCACTATAAGGTGA 421
   |||
QY    217  GGTTCCCTATGACGCTATCATGAGAAATTAATCATGAAAGATCCAACTATAAGCT 276
   |||
Db    422  AGTTCTTTGATGCGCTCTTTAGTGAAGAGCTATTAATGAAAGCCCACTATCAGCC 481
   |||
QY    277  AAAAGATGAGATATTTGTAATGAGGTCAAGGCTGATATGTTATCAAGTATGATGANA 336
   |||
Db    482  GAAAGATCAAGATATCGTACAGTTAGTTAAGGAGATATATTATTAAGTATGATGANA 541
   |||
QY    337  ATACTATGTTTACCTTAAGATGCTGCCAGCGGATTAACGTCCTGACAAAAGAGAAAT 396
   |||
Db    542  GTATTATGTTTACCTTAAGATGCTGCCAGCGGATTAACGTCCTGACAAAAGATGAAAT 601
   |||
QY    397  CAATCGACAAAAACAAGAGCATGTCATCATGAGGTGAGTCCCAAGAAACGATG 456
   |||
Db    602  CAATGCGCAAAAACAAGAGCATGTCATGAGGTGAGTCCCAAGAAACGATG 650
   |||
QY    457  TCGTGTTCCTGCGACGTTGCGCAAGAGCGCTATATCAAGATGATGTTATCTTTAA 516
   |||
Db    651  GAAGTATCTGTAGCAAAATTAACAGGAGCGTTATCTACCATGATGTTATATATAG 709
   |||
QY    517  TGCTTCTGATATCATAGAGATCTGCT 544
   |||
Db    710  AGTTGAATTTAGAAATGACATCATAGT 737
   |||

RESULT 42
LOCUS      AR219026      1684 bp      DNA      linear      PAT 25-SEP-2002
DEFINITION Sequence 258 from patent US 6420135.
ACCESSION  AR219026
VERSION    AR219026.1 GI:23319960
KEYWORDS
SOURCE     unknown.
ORGANISM   Unclassified.
REFERENCE  1 (bases 1 to 1684)
AUTHORS   Kunisch, C.A., Choi, G.H., Dillon, P.S., Rosen, C.A., Barash, S.C.,
          Fannon, M.R., and Dougherty, B.A.
          Streptococcus pneumoniae polynucleotides and sequences
          Patent: US 6420135-A 258 16-JUL-2002;
          Location/Qualifiers
            source      /organism="unknown"
BASE COUNT      593 a      292 c      345 g      453 t      1 others
ORIGIN
Query Match      10.2%; Score 243.4; DB 6; Length 1684;
Best Local Similarity 60.3%; Pred. No. 1.7e-41;
Matches 493; Conservative 0; Mismatches 236; Indels 88; Gaps 2;

```

```

QY 1629 ATGCATATGTAACGCTCATATGCGCCATAGTCACTGGATTGGAAAAAGATAGCCTTTCTG 1688
DB 1 ATGCCTATGTAATCTCCACATATAGACCCATAGCCATGGATTAAAAAAGATGTTTGTCTG 60
QY 1689 ATAAAGAAAAAGTTGCGAGCTCAAGCCTTAATAAAGAAAAAGTATCTTACCTCATCTC 1748
DB 61 AAGCTGAGAGAGCGGCA -CCGAGGCTTAATGCTAAAGAAAAAGTTTGAACCCCTCTTGA 119
QY 1749 CAGACGCAATGTTTAAAGCAATCCAACTGGAGATAGTGCACGACTTTTAAATCTGTG 1808
DB 120 CAGACCATCAGATTTCAGAAATCTGAGGCAAAAGGCGAAGAACCTTAATCAACACCGG 179
QY 1809 TGAAGGGGAAAAAGCAATTCCTGCTGCTGCACTTCCATATATGTTGAGACATAGTTCG 1868
DB 180 TGAAGGCGCTTAAGAAAGTGCACCTTGATCGATGCTTACAACTTCAATATCTGTAG 239
QY 1869 AGGTTAAAAAGGTAATTTGATTTCTCTCATAGGATCATTTACCAATAATAAATTTG 1928
DB 240 AAGTCAAAAAAGGTAGTTTAATCATACCTCATTTATGACCATTAACATCAAAATTTG 299
QY 1929 CTGTGTTGATGATACACATACAAAGCTCCAAATGGCTATACCTTGGAAATTTGTTG 1988
DB 300 AGTGTGTTGACGAAAGCCTTTATGAGGCACTTAAGGGGTATACCTTGAAGATCTTTTG 359
QY 1989 CGACGATTAAGTACTACGTAGAACACCTGACGAAAGTCCACATTTCTAATGATGATGG 2048
DB 360 CGAGCTGTCAATGACTATGCTCAACATCCAAAGAACGTCGCTTCAATATATGTTTGG 419
QY 2049 GCAATGCCAGTGAAGATGTTTGAAGCAAAAGCAACAGTGTGATGATCCAAATAAGAACT 2108
DB 420 GTAACGCTAGCGACCATGTTCAAGAAACAAATAATGTCAAGCTGATACCAATCAACGG 479
QY 2109 TCAAGCGGATGA----- 2121
DB 480 AAAAAACAAGGAGAGAAACCTCAGACAAAAAAGCTGAGAGAAACCCCTCGAGAG 539
QY 2122 -----AGAGCAGTAGAGGAAACAC 2141
DB 540 AGAAACCCGAAAGGAGAAACAGAGTCTCCAAAAACAAGAGAACCAAGAAATAC 599
QY 2142 CTGCTGAGCCAGAAAGTCCCTCAAGTAGAGTGAAGTGAAGAAAGCCCACTCAAGAG 2201
DB 600 CAGAGGATCAGAAAGCCTCAGGTCGAGACTGAAGAAAGTTGAAGAAACCTGAGAGG 659
QY 2202 CAGAAATTTGCTTCCGAAAGTGAAGGATTTCTAGTCTGAAGCCCAATGACAGAAATC 2261
DB 660 CTGAAGATTTACTTGAAAAATCCAGATCCAAATATCAAGTCCAAAGGAGACTC 719
QY 2262 TAGCTGTTTGAAGAAATATTTGAATCTTCAAAATATGATTAACAATAGTATCAATGCGAG 2321
DB 720 TCACAGGATTAATAAATAATTTACTATTGTCACCCAGGACAAACATCTATTATGCGAG 779
QY 2322 AAGCAGAAAAATTAATCTGCTGTTTAAAGAGATTA 2358
DB 780 AAGCTGAAAAATTAATGCTTTATTAAAGAGAGATTA 816

```

```

RESULT 43
LOCUS BD003938 1684 bp DNA linear PAT 31-JAN-2002
DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.
ACCESSION BD003938
VERSION BD003938.1 GI:18631899
KEYWORDS JP 2001501833-A/258.
SOURCE unclassified
ORGANISM unclassified
REFERENCE 1 (bases 1 to 1684)
AUTHORS Kunsch, C.A., Choi, G.H., Dillon, P.J., Rosen, C.A., Bara, S.C.,
TITLE Polynucleotide of Streptococcus pneumoniae and sequence
JOURNAL Patent: JP 2001501833-A 258 13-FEB-2001;

```

```

COMMENT HUMAN GENOME SCIENCES INC
OS Unidentified
PN JP 2001501833-A/258
PD 13-FEB-2001
PF 30-OCT-1997 JP 1998520718
PR 31-OCT-1996 US 60/029960
PI CHARLES A KUNSCH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, PI
STEVEN C BARASH,
MICHAEL FANNON, BRIAN A DOUGHERTY
PC C12N15/09, A01K67/027, C07K14/315, C07K16/12, C12N1/15, C12N1/19,
PC C12N1/21,
PC C12N5/10, C12P21/02, C12Q1/68, G06F17/30, C12N5/00, C12N5/00, PC
G06F15/40
CC Strandedness: Double;
CC Topology: Linear;
FH Key 1.1684 Location/Qualifiers
FT source 1.1684 'Unidentified'.

BASE COUNT 593 a 292 c 345 g 453 t 1 others
ORIGIN

Query Match 10.2% Score 243.4; DB 6; Length 1684;
Best Local Similarity 60.3%; Pred. No. 1.7e-41;
Matches 493; Conservative 0; Mismatches 236; Indels 88; Gaps 2;

QY 1629 ATGCATATGTAACGCTCATATGCGCCATAGTCACTGGATTGGAAAAAGATAGCCTTTCTG 1688
DB 1 ATGCCTATGTAATCTCCACATATAGACCCATAGCCATGGATTAAAAAAGATGTTTGTCTG 60
QY 1689 ATAAAGAAAAAGTTGCGAGCTCAAGCCTTAATAAAGAAAAAGTATCTTACCTCATCTC 1748
DB 61 AAGCTGAGAGAGCGGCA -CCGAGGCTTAATGCTAAAGAAAAAGTTTGAACCCCTCTTGA 119
QY 1749 CAGACGCAATGTTTAAAGCAATCCAACTGGAGATAGTGCACGACTTTTAAATCTGTG 1808
DB 120 CAGACCATCAGATTTCAGAAATCTGAGGCAAAAGGCGAAGAACCTTAATCAACACCGG 179
QY 1809 TGAAGGGGAAAAAGCAATTCCTGCTGCTGCACTTCCATATATGTTGAGACATAGTTCG 1868
DB 180 TGAAGGCGCTTAAGAAAGTGCACCTTGATCGATGCTTCAATATATGTTTGG 239
QY 1869 AGGTTAAAAAGGTAATTTGATTTCTCTCATAGGATCATTTACCAATAATAAATTTG 1928
DB 240 AAGTCAAAAAAGGTAGTTTAATCATACCTCATTTATGACCATTAACAATCAAAATTTG 299
QY 1929 CTGTGTTGATGATACACATACAAAGCTCCAAATGGCTATACCTTGGAAATTTGTTG 1988
DB 300 AGTGTGTTGACGAAAGCCTTTATGAGGCACTTAAGGGGTATACCTTGAAGATCTTTTG 359
QY 1989 CGACGATTAAGTACTACGTAGAACACCTGACGAAAGTCCACATTTCTAATGATGATGG 2048
DB 360 CGAGCTGTCAATGACTATGCTCAACATCCAAAGAACGTCGCTTCAATATATGTTTGG 419
QY 2049 GCAATGCCAGTGAAGATGTTTGAAGCAAAAGCAACAGTGTGATGATCCAAATAAGAACT 2108
DB 420 GTAACGCTAGCGACCATGTTCAAGAAACAAATAATGTCAAGCTGATACCAATCAACGG 479
QY 2109 TCAAGCGGATGA----- 2121
DB 480 AAAAAACAAGGAGAGAAACCTCAGACAAAAAAGCTGAGAAAAACCCCTCGAGAG 539
QY 2122 -----AGAGCAGTAGAGGAAACAC 2141
DB 540 AGAAACCCGAAAGGAGAAACAGAGTCTCCAAAAACAAGAGAACCAAGAAATAC 599
QY 2142 CTGCTGAGCCAGAAAGTCCCTCAAGTAGAGTGAAGTGAAGAAAGCCCACTCAAGAG 2201
DB 600 CAGAGGATCAGAAAGCCTCAGGTCGAGACTGAAGAAAGTTGAAGAAAACTGAGAGAG 659

```

QY 2202 CAGAGTTTGGCTTGCAGAAAGTAACGATTTAGTCTGAAAGCCATGCAAGAAACTC 2261
 DB 660 CTGAAGATTACTTGGAAAAATCCAGATCCAAATTATCAAGTCCAAATGCAAGAGACTC 719
 QY 2262 TAGTGCTTTACGAATAATTTGACTCTTCAAAATTATGATAACAATAGTATCATGCGAG 2321
 DB 720 TCACAGGATTAATAAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 779
 QY 2322 AAGCAAAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 2358
 DB 780 AAGCTGAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 816
 RESULT 44
 AX602204/C
 LOCUS AX602204 75248 bp DNA linear PAT 17-FEB-2003
 DEFINITION Sequence 133 from Patent WO02092818.
 ACCESSION AX602204
 VERSION AX602204.1 GI:28402073
 KEYWORDS
 SOURCE Streptococcus agalactiae
 ORGANISM Streptococcus agalactiae
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 REFERENCE 1
 Glaeser, P., Rusniok, C., Chevallier, F., Frangeul, L., Laliou, L.,
 Zouine, M., Couve, E., Buchrieser, C., Foyat, C., Tiliou-Cloc, P. and
 Kunst, F.
 Streptococcus agalactiae genome sequence, use for developing
 vaccines, diagnostic tools, and for identifying therapeutic targets
 Patent: WO 02092818-A 133 21-NOV-2002;
 INSTITUT PASTEUR (FR) ; CENTRE NATIONAL DE LA RECHERCHE
 SCIENTIFIQUE (CNRS) (FR)
 FEATURES
 source location/Qualifiers
 1..75248
 /organism="Streptococcus agalactiae"
 /mol_type="genomic DNA"
 /db_xref="taxon:1311"
 BASE COUNT 24346 a 14447 c 11608 g 24847 t
 ORIGIN
 Query Match 9.2%; Score 219.6; DB 6; Length 75248;
 Best Local Similarity 48.0%; Pred. No. 1.9e-36;
 Matches 1083; Conservative 0; Mismatches 1045; Indels 127; Gaps 11;

QY 461 -----GTTGCTTGGCAGGTTGCAAGACGCTATTAACAAGATGATGTTATATCT 512
 DB 22401 CAGCAGGCTACGTAAACGAAAGCTAGCGGTGCTATTAACATGATATGTTATGTTT 22342
 QY 513 TTAATGCTTGTATATCATAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 572
 DB 22341 TTAGTCCAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 22282
 QY 573 ATTACATTTACATTTCTTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 630
 DB 22281 ATTTCATTAATACCGAAGCTGATTTATCTCCATCAGAGTATACAGCCGACAGGACT 22222
 QY 631 -----CTTCTATCTGGGCGAGAAATCTGCAATTCGAACACT 671
 DB 22221 ATTGGAACCGCAAGCTGGCGGATAGGTAATAGTCTCAAAACCATCAATTCAGATTCT 22162
 QY 672 ATGCGCGCAAAATAGCATTAACATTCAGAAACCACTGGTACCTTCT----- 721
 DB 22161 ACATTCATGCAAGTGCACCAAGCGGTATGATTAACAAGAACATGCTATATGCCCCCTA 22102
 QY 722 -----GTAAACATCCAGAACTACAAATA 746
 DB 22101 TTTCTATTCCAAAGTCACTACGCAATCACTGTCAAAACCAAGCTGATATCATGCTA 22042
 QY 747 CTAAACAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 806
 DB 22041 CGGACCAAAACCATGACACTTACACCAACCAATCAACCAATTCCTGTTAGACA 21982
 QY 807 GTCTCTTGAACAGCTCTTACAACTGCTTGAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 866
 DB 21981 AAATGTTGAACGTTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 21922
 QY 867 TTGCTTTGATTCAGACCAAAATCAAGTCCAGACGCTAGAGGTGTCAGTCCACACG 926
 DB 21921 TGGTTTATGACCCAGCAAGTAATGCTTTCACAGCAATGCTGTTTCTATTCGACG 21862
 QY 927 GAGATCATTTACACTTCACTTCCCTTCACTCAATGTCGAATTTGGAAGCAAGTGCCTC 986
 DB 21861 GTAAATCATTTTCACTTTTATTAATTAAGATATCTTCCATTAAGATTGAAGCAACAA 21802
 QY 987 GTATTATTCCTCTGTTATGCTTCAACCAATGGGTACAGATTCAGAGCCAGAACAC 1046
 DB 21801 GGATGCTGGCAGAGC---ATAGAGCAATCATATATGATGCAATTAAGGAAAAAGTTCTA 21745
 QY 1047 CAAGTCCAAACCGACTCCGAACTAGTCCAGGCCGCAACTGACCAAAATCTTAA 1106
 DB 21744 CAGAGAAACCAAGCATATTTCTCATGAACCTAATPAAGAACTCAGACAGAGAAAC 21685
 QY 1107 TAGACTCAATTTCTTTGTTGTTAGTCAAGTGTGTAAGAAAGTGGGAAGATATGAT 1166
 DB 21684 ACCATGCAAGTACAC---GAAGACCAAGCTTAAGGCAACCAATAGCCAGATGTC 21629
 QY 1167 TCGAAGAAAAGGCGCTCTCTGTTATGCTTTGCGAAAGATTTACATCTGAACTGTTA 1226
 DB 21628 TACAGTGTCAAGAAATTTGAAGAGCAAAAAGCTGCTAATATACCAATCTGATGTT 21569
 QY 1227 AAATCTTGAAGCAAGTTATCAAAAAGAGAGTTCACACACTTAATCTGCTA 1286
 DB 21568 ACATTTTGAAGCCAAAGATATTAAGAAAGTACAGTACAGTATGCTATCCACATA 21509
 QY 1287 AAGAAATGTTGCTCTGCTGTAAGCAAGATTTTAAGTAAGCAATATATCTGTTA 1346
 DB 21508 TGACACATG-----AGCATTTGGTACCAAGAAAGATTTATCAGAGTCCGAAATTAAG 21455
 QY 1347 AGGCTCATAAAGCTTGTGTTGAAATTAAGGCTGTAATTTGATTTCCAGGCTTAGACA 1406
 DB 21454 CAGCTCAAGAAATTTCTTCAAGAAATCTGAAGCAAAATCAAGACAAACCAAAAAGAGTA 21395
 QY 1407 AATTAATGAACGCTTGAATGA-----TGAATGACTAATPAAGAAATTTGTAAGT 1460
 DB 21394 AAACAGCTCAAGAAATCTATGAGGCAATTTGAACCAAGCAATTTGTAAGCTGAAGATT 21335

QY 1461 ATTATTTGGCATTCTCTAGACCAATTACCCATCCAGACG---ACTTGGCAACCAATT 1517
 Db 21334 TATTATTTGGATTGACAAAGGACAGACTATTAACATATGATACATTCTTATCTCTATA 21275
 QY 1518 CTCAAATTGATTAATCTGAAGCGAAGTTCTGATTGCTCAATTAGCTGATAGATACAA 1577
 Db 21274 AAGATCATTTACCATATATGTGAAATTAATGTTTGAAGAAAGAAATATCTTTAGCTG 21215
 QY 1578 CGTCAGATGGTTACTTTTGTATGACATGATATATAGATGATGAAGAGATGATATG 1637
 Db 21214 ATTCAAGATAAGACATATATCTTTAGAGACTATTTAGCTACGCTAAATATTCATGATGC 21155
 QY 1638 TAAAGCCTCATATAGGCGCATAGTCACTGATTTGAAAAAGATAGCCTTTCTGATAGAAA 1697
 Db 21154 ACCAGAAAAAGCTCTAAAGTTGAAGATGGGGTAAAGATGCTGAATTTATAGAGAA 21095
 QY 1698 AAGTTGCAGCTCAGCCTTATATCTAAAGAAAAAGTATCTTACCTTCATCTCCAGACGAG 1757
 Db 21094 AGGACTCTAATAAAGACAGATA-----AACCAAGTCTGCACCAACTGATATAATCA 21041
 QY 1758 ATGTTAAGCAAAATCCAACTGGAGATAGTGCAGACGATTTTCAATGCTGMAAGGG 1817
 Db 21040 CATCAAAATTCATGATGACAAAACTTAAGTGTGCAGAAAGTATTCAAACAGCAAAACAG 20981
 QY 1818 AAAAAGCAATTCACCTCGTTCGATATATGTTGAGCATACAGTTGAGTTAA 1877
 Db 20980 AAAAAATTGACCGTTTATATAAATTGCTGCTCACATGSCATATGCACTGATTTGAG 20921
 QY 1878 ACGGTAATTTGATTAATCTCTAATAAGATCATTTACCATATAATTAATTTGCTGGTTG 1937
 Db 20920 ATGATCAATGATGTTCTCTCATCATGATCATTAATGATGCTTATGATGATGCTTTG 20861
 QY 1938 ---ATGATCACACATACAAAGCTCCAAATGGCTATACCTTGGAGATTTGTTGGAAGA 1994
 Db 20860 ACAAAGGGTGTATTAGAAAGACCAAGAGGCTATACATTCACACAACTCTTCTCAACA 20801
 QY 1995 TTAAGTACTAGTAGAACACCTCGACGACGCTCCATTTCAATGATGATGGGCAATG 2054
 Db 20800 TTAATATCTACATGSAACATCTTAATGATTAACAAA---AGAAAAAGTTGGGACACG 20744
 QY 2055 CCAAGTACATGTTGTTAGCAAGAAAGCAACAGTGAAGATCCAAATTAAGACTTCAAG 2114
 Db 20743 ACAGATGATCATTAACAAAGGCTCAATTAAGACAAATAAAGCCAAATATATGCTCCAGATG 20684
 QY 2115 CGATGAAGAGCCATAGAGAAACCTGCTGACCCAGAAAGTCCCTCAAGTAGAGACTG 2174
 Db 20683 AAGAACTGAAGATTCAAGGAAAGTAACTCACTCACTATGTTTATGATGTTAATAAG 20624
 QY 2175 AAAAAGTAGAAGCCCACTCAAAAGAGCAAGTT 2209
 Db 20623 GTTCAGACGAAGAAACCAAGAAAAACAAGAAAT 20589

RESULT 45
 SAG766854/c 98050 bp DNA linear BCT 04-NOV-2002
 LOCUS Streptococcus agalactiae NEM316 complete genome, segment 12.
 DEFINITION AL766854 AL732636
 ACCESSION AL766854.1 GI:24413438
 VERSION
 KEYWORDS
 SOURCE Streptococcus agalactiae NEM316
 ORGANISM Streptococcus agalactiae NEM316
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.

REFERENCE
 1 Glaser, P., Rusniok, C., Chevallier, F., Buchrieser, C., Frangeul, L.,
 Zouine, M., Couve, E., Lailouli, L., Msadek, T., Poyart, C.,
 Trieu-Chuc, P. and Kunst, F.
 Genome sequence of Streptococcus agalactiae, a pathogen causing
 invasive neonatal disease
 JOURNAL Mol. Microbiol. 45 (6), 1499-1513 (2002)
 MEDLINE 22242508
 PUBMED 12354221

REFERENCE 2
 AUTHORS Glaser, P., Rusniok, C. and Frangeul, L.
 TITLE Direct Submission
 JOURNAL Submitted (31-MAY-2002) Glaser P., Institut Pasteur, Genomique des
 Microorganismes Pathogenes, 25, rue du Docteur Roux, 75724 Paris
 Cedex 15, FRANCE. E-mail: pglaser@pasteur.fr Phone: +33 (0)1 45 68
 89 96, Fax: +33 (0)1 45 68 87 86
 FEATURES
 source
 1. 98050
 /organism="Streptococcus agalactiae NEM316"
 /mol_type="genomic DNA"
 /strain="NEM316"
 /serovar="serovar III"
 /db_xref="taxon:211110"
 101. 410
 /gene="gbs1858"
 101. 106
 /gene="gbs1858"
 117. 410
 /gene="gbs1858"
 /codon_start=1
 /trans_table=1
 /product="Unknown"
 /protein_id="CAD47517.1"
 /db_xref="GI:24413439"
 /translation="MILINILALPICFLVELSAGILKLEMHKOHGYYPLERLMTGN
 ILTVSLVLTHTLIHNINSLPLSITVVLISKSLISLISYLLKRSKQKV"
 complement (405. 1725)
 /gene="gbs1858"
 complement (405. 424)
 /gene="gbs1858"
 /gene="gbs1858"
 complement (433. 1725)
 /note="Similar to adenylosuccinate synthase"
 /codon_start=1
 /trans_table=1
 /product="Unknown"
 /protein_id="CAD47518.1"
 /db_xref="GI:24413440"
 /translation="MTSVVVGQMGDEGKKTIDPLSADAEVIARYOGGDMAGTIV
 IDNKKFRLHLPISGIFFEKISVIYGNVVPVSLVLELVLHGEVTTDLRLISDA
 HVLIPYHKLDOLEADKGNKIGTITKIGPVMYDKAAVGRIALDLBREYAEEL
 KINIAERNRLFEKGYVDSTPLEPDIPEEYVEYGOQIKQYVTLDSVILNDLADGRVL
 FEGAGYMLIDOGTYPFTSSNPVAGVTIGSGVGSKINKVVGCKAYTSRGDGP
 FPEELPDEVGDRIRLEIGKEYGTTGPRPRVGVGDSVYMRHSRRVSGITNLSINDVL
 SGLDVTICVAYLDKRDIDYPPASLEQLRCKRPIYELPGRMSBTDIACNSLDLLEN
 ARNVRVGVGLVGRISTFSVGPGRBOTNILESWSNI"
 complement (1983. 3130)
 /gene="gbs1860"
 complement (1983. 2012)
 /gene="gbs1860"
 complement (2096. 3118)
 /gene="gbs1860"
 /note="Similar to transcriptional regulators and to PTS
 enzyme II"
 /codon_start=1
 /trans_table=1
 /product="Unknown"
 /protein_id="CAD47519.1"
 /db_xref="GI:24413441"
 /translation="MNIIIGTSLILVLAIFLFFNYKAPYTKMGAALASAAACPLV
 EAFODSPFGVILGFOPFSEVGGANGSLGVAAIIVLAIGTGVAVLIGLSVGTG
 IITGFLAGYVGLPYKMEBNIRGGLDISITIIIGAPLRLVAKLLPLINSTLTITG
 DILTSGAHSPILMGIILGGIIVVATAPLSSNAUTMLGLTGKPMALGALSVGGSF
 MNGVLPHKLKLSRKQNIARAVERLTQADVTSANPIPIVYTNFVGACGILLAMKL
 VNITPGPATIAGAAVFAVNPMLKVLITAGCIIISLLAGYFGGIVFKDYKLVTKBE
 LQARDN"
 complement (3125. 3130)
 /gene="gbs1860"
 3465. 4209
 /gene="gbs1861"
 3465. 3470
 /gene="gbs1861"

RBS
 gene
 RBS

```

CDS
3480..4154
/gene="gbs1861"
/note="Similar to unknown proteins"
/codon_start=1
/transl_table=11
/product="Unknown"
/protein_id="CAD47520.1"
/db_xref="GI:24413442"
/translation="MDKEKLDYWKTIITFLHNVLDGNYEVLHVVDENDIYIGELVNS
HISGRITSSPLTTFALDIKNVYKEDPVTYKAYASPLNKRGSTPFIKAONEL
EEMLCINLDISAONIALDITLVNLNKKIIPKSPQKISLPQOEPPVLSGNIDDI
ISEIVDPSLLNQNILHSOEKVEIVSKLHEKGVFLKGAVSKVAEVLNISEPSVRYL
KKIEFH"
terminator
4188..4209
/gene="gbs1861"
4309..6608
/gene="gbs1862"
4309..4314
/gene="gbs1862"
4324..6576
/gene="gbs1862"
/note="Similar to gamma-glutamylcysteine synthetase (for
the N-terminal part) and to cyanophycin synthetase (for
C-terminal part)"
/codon_start=1
/transl_table=11
/product="Unknown"
/protein_id="CAD47521.1"
/db_xref="GI:24413443"
/translation="MIDRLQSHSLPLDQATFGLERESLRHQPTORVAQTPHPK
TIGSRNHYHYIQDYSEPOLIELITPIAKSOEAIKPLKASIDVAGRSINHDELYMPLS
MPKYREDIOIAQLEDAFEYDYRYKLEKTYGLIOSIGIHNLGLGQELTJPEL
SOADNADRONLYMKLSONPLRYBWLTVLYGASPAEEDPLDOKLNPVSLRSHR
LCVYVHKDRISTYSKYDYNLDENAVKSGQILAEKEFSPVRLRSKRCRYNLEKCI
TUEPRTFDLPSPSPIGITOEVDIVHLFLALMLDSSSHIDQDIKEANRLNDLAL
SHPEKLPNQAPVSDLVADMQSVLOHFNLSPPYQDLLESVKQIOSEPILTVAGOLEM
IGLSLETFGORGQGIYHDYAMEAPYALKGVEYMETSTOLLFPDVIQKVEVEVLEQ
DOFLKLNHSHIEYVNGNMTSKDNYIVPLAMANKVTKKIDEXHEPFPDEDFDR
KEALNFYSQIDOKPIVYKPKSTNFGISIFPKTSANLASYEKADIDAFEDSAIIEE
YIEGTEYRFYVEGDCIATVLRYAANVVDGHIITTSQVYKLNQNPRLQYDRSPLEV
IELGEVEQMLEQGGITVNSIPPEGTKIELRNSNISTGDSIDVNTNMDPYKQILAA
EVAEMGAVVCGVDLIIPNATQAYSXDKKNATCIELNFMPLMYMTYCOEGFGOSITP
RIIAKLFPEL"
terminator
6588..6608
/gene="gbs1862"
6700..7518
/gene="gbs1863"
6700..7518
/gene="gbs1863"
/note="Similar to unknown proteins"
/codon_start=1
/transl_table=11
/product="Unknown"
/protein_id="CAD47522.1"
/db_xref="GI:24413444"
/translation="WYKRYKSNSHLITLIGIVLASLSDPFWSVIATLLIGGY
QVYVYDKRKLKRTFSNQRILEAKSDIKETDDDIRLLELKDNRSKEITKIAHQILP
QULIRNEANQIQKALEBNITKRTKANATFSEINIEQILKIHASELSEPIDQDEM
IRIAPELKFEYNIQODHFAILKIKIEANRKAELAIHQANKRFPDVIAGYIRIKOS
PNFNNAKERLEQALQAIKKFNLDIDETLRQINESDKDFDVLRMQDERNSK"
7504..8821
/gene="gbs1864"
7504..7509
/gene="gbs1864"
7515..8771
/gene="gbs1864"
/note="Similar to other proteins"
/codon_start=1
/transl_table=11
/product="Unknown"
/protein_id="CAD47523.1"
/db_xref="GI:24413445"
/translation="MSDTFNPDIQIADNVAITKTDKTEIISNQTTSTQTAFAPEKL

```

```

TPAQKSAISEKTPALVDTFVGDONALLDFGASVEGVNTVNHILISEQKIOIPQVD
ILKNANREINNGFIAYKDATPALEKRPULIOKLKROKSTLSOEFPDSONEOKMDI
MAANVYKQEDTLIARNTIVSAEMIIEDNTKSIENLVIVATIESOAEANRAHIOEI
IALDSQTSYQIKSNQOLARMTVINLTLEQOHTREYVSRIVANATTIPQKNLVKSDM
ROKLGMRNNTIIPYKLSIAQGMQOQSVKSVTADAVIANNANALOMLAETSKAIP
MLEKTAQSPVTSIKSVTALAEISLVAAQNGNIIIAIDKGRKERAQLESVAKSAETINDS
VKIRDKIYEAELNLEKSTQEKVDES"
8799..8821
terminator
/gene="gbs1864"
8911..9786
/gene="gbs1865"
8911..9786
/gene="gbs1865"
/note="Similar to other proteins"
/codon_start=1
/transl_table=11
/product="Unknown"
/protein_id="CAD47524.1"
/db_xref="GI:24413446"
/translation="MDKIIKSTSGSPRAYVLDCTETRYTAQKHQTLSSSTVALGR
TLIANQIIAANKGSKSVTVKVTGDDSSREHIIISVADTKNVGYIQNTQVDIKKATG
EVLVDFPENGHEFVITDYATQPYTSTTPIITGIEGDFAYLTSESQTPSAGVLNV
LLDDEKXKAVAGFMLOVLPASDEISRYEKRIQEMPSISLSLENNIEBLSLAIX
GEDDYKRLSEDSLAFYCDOSKRFEPALALTIGTKELQAMKDEKVEITICPCNQTY
FTEEDLEKIINDSIK"
9755..10747
/gene="gbs1866"
9755..9760
/gene="gbs1866"
9770..10747
/gene="gbs1866"
/note="Similar to transcription regulator"
/codon_start=1
/transl_table=11
/product="Unknown"
/protein_id="CAD47525.1"
Query Match
9.2%; Score 219.6; DB 1; Length 98050;
Best Local Similarity 48.0%; Pred. No. 1,9e-36;
Matches 1083; Conservative 0; Mismatches 1045; Indels 127; Gaps 11;
QY 60 ATATAGATGGAAGAAAGCGAGCGCAAAACCGAAGATTGACTCCTGATGAGTTAGCA 119
DB 77948 ATTAATTCGATATGCAAGACACAGCAACGCTGTAAATAACGCCCGCAACATAGTA 77889
QY 120 AGCGTGAAGAAATCAATGCTGAGCAAAATGTCATCAAGATTAACAGACCAAGGCTATGTCA 179
DB 77888 AAGAAGAGGCAATATCTGAGCAACAAATGCTTAAATTACTGACGACGTTATGTGA 77829
QY 180 CTTCAATGCGGACCACTATCTATTATCAATGTAAGGTTCCTTATGACGTATCATCA 239
DB 77828 CTTTCATGAGACCACTTACATTAATTCMAATGGGAAAGTGCATATGATGCAATCATTA 77769
QY 240 GTGAAGAAATTAATCTGAAAGATCCAACTATAGCTAAAGATGAGATATTGTTAATG 299
DB 77768 GTGAAGAAATTAATGAAAGACCTAGCTATGCTTTATATAAGTATGTCATTAATG 77709
QY 300 AGGTCAAGGGTGAATATGTTATCAAGGATAGATGAAATATCTATGTTTAACTTAAGANG 359
DB 77708 AAGTCAAGATGTTATATATATCAAGGTTAAGGTAAGTAATTAACCTATCTTAAGAAG 77649
QY 360 CTGCCACCGCGATTAACGTCCTTACAAAAGAGAAATCAATGACAAAAACAAGACATA 419
DB 77648 GTAGCAACGAAACCAATGTTCTTACAAAAGAAACCAATCCAAAAGCAACCGGAAGATGT 77589
QY 420 GTCAACATCGTGAAGTGAATCTCCAAAGAAAGATGTGCT----- 460
DB 77588 CTAAGGTGCTTCTAAGGAGATACGAAAGCATAGTAGTCAAAAACTCAAGCTTAT 77529
QY 461 -----GTGECCTTGGCAGCTTGGCAAGACGCTATATCTACAGATGATGTTATCT 512
DB 77528 CAGCCAGGTACGTAGAGCAAGCAAGCTAGCGGTGCTATATCTACAGATGATGTTATGTTT 77469

```

```

Oy 513 TTAATGCTTGTATATCATAGAGATAGTGTATGCTTATATCTCTCATAGAGATC 572
Db 77468 TTAGTCCAAAGATCTCATGTATGATGATGGAGATGCTTTTATGATACATGCGATC 77469
Oy 573 ATTACATTTACATCTTCAAGAAATGATATCATGAGAGTTGGCTCTGAGAAAGC-- 630
Db 77408 ATTTCATTTATATCCGAAAGCTGATTTATCTCATAGAGATTATCACCGCAGAGGAT 77349
Oy 631 -----CTTCATATGCTGTCAGAGAAATCTGCAAAATTCAGAAACT 671
Db 77348 ATTGGAACCGAAGACTGGCCGATCAGTAAGTCAAAACCAATCAATCAATGATAT 77289
Oy 672 ATCCGCGCAAAATATGCAATTAACCTTAAGAAACAATGCGGTACTCTT----- 721
Db 77288 ACATTCATGCAAGTCCACCAAGCGGTATATGATATCAACAGAAAGATGCTATGCCCCA 77229
Oy 722 -----GTAGCAATCCAGAACTACAAATA 746
Db 77228 TTTCTATTCAGAGATCACTCAGCAAAATCACTGGTCAAAACCACTGGTATATGCTA 77169
Oy 747 CTAAACAAGCAACAAGCAAACTAATCAAGTCAAGCAAGTCAAGTATGATGATGATA 806
Db 77168 CGGACCAAAACACCATGCACTTACCAACCAATCAACAAAGATTTCTGTTAGACA 77109
Oy 807 GTCTCTGAAACAGCTCTACAACATGCTTGGAGTCAAGCAATGATGATGATGCGC 866
Db 77108 AAATGTTAAAGCTTATATGCTCAACCGTTATATGCTGCTCAGTAAAGATGATG 77049
Oy 867 TTGCTTTGATTCAGCAACAATCACAAGTCCGAACGCTAGAGGTGTTGCAAGTCCACAG 926
Db 77048 TGGTTATGACCCAGCAACAAGTTATGCTTTCACAGCAATCGTCTTCTATTCGCGACG 76989
Oy 927 GAGATCATTCACATCTCCTTCTCTCAAAATGCTGAAATGGAAGAAGAAATCGCTC 986
Db 76988 GTAATCATTTCTCATTTATCTCATTAAGATATATCTTCATTAAGATTAGAACAA 76929
Oy 987 GTATATATCCCTTGTATCTGTTAAACCATTTGGGTACCAAGATTCAGGCGAGAAAC 1046
Db 76928 GGATGTTGGCAGAGC---ATAGAGACATCATATTTGATGATTTAGGAAAAAGTTCTA 76872
Oy 1047 CAAGTCCAAACCGACTCCGGAACCTAGTCCAGGCGCCGCAACTGACCAAAATCTTAAA 1106
Db 76871 CAGAGAAACCAAGCATATTTCTCATGAACCTAATAGGAACCTCACAAGAGAAAGAC 76812
Oy 1107 TAGACTCAATTTCTTTGTTAGTCAAGCTGTAAGAAAGTGGGAAGATATGAT 1166
Db 76811 ACCATGCAAGTAAACCC---GAAGAACAACGTAAAGGCAACCAATATGCCAGATTGTC 76756
Oy 1167 TCGAAGAAAGGCACTCTCTGTTATGTTTTCGAAAGATTTACATCTGAAACTGTTA 1226
Db 76755 TACAGTGTCTCAAGAAATTTGAAGAGCAAAAAGCTGTAAATACCAACATCTGATGTT 76696
Oy 1227 AAAATCTTGAAGCAAGTTATCAAAAACAAGAGTGTTCACACACTTAACTGCTAAA 1286
Db 76695 ACATTTTGTATGCCAAAGATATTTAAAAAAGATACAGTATGATTCATTCACATA 76636
Oy 1287 AAGAAATGTTGCTCTCTGTCGCAAGATTTTATGATTAAGCAATATATCTGTTAACTG 1346
Db 76635 TGACACATG-----AGCATTTGGGTACCAAGAAAGATTTTACAGATGCGAATTTAAAG 76582
Oy 1347 AGGCTCATAAAGCCTTGTGTTGNAATTAAGGTCGTAATTTCTGATTTCCAGCCTTAGACA 1406
Db 76581 CAGCTCAAGAAATTTCTTTCAGAAATATCTGAAGCAATTCAGAAACCAAAACCAAGGTA 76522
Oy 1407 AATTATAGAAAGCTTGAATGA-----TGAATGCACTAATTAAGAAAAATTTGTAATG 1460
Db 76521 AAACAGCTCAAGAAATCTATGAGCAATTTGAACCAAAAGCAATTTGTTAACTGTAAGTT 76462
Oy 1461 ATTATTTGCAATTTCTAGACCAATTTACCATTCAGAGCG---ACTTGGCAACCAAAAT 1517
Db 76461 TATTTATTTGGAATTCACAAAGGACACTAATTAAGAAATGTAATTTGTAATTTCTCATTA 76402
Oy 1518 CTCAATTTGAGTATCTGAAAGCAAGTTGTAATTTGCTCAATTTAGCTGTAAGTATACAA 1577

```

```

Db 76401 AAGATCATACCATATATGGAATTTAAATGTTTGAATGAAGAAAAAGATCTTTAGCTG 76342
Oy 1578 CGTCAATGTTTACATTTTATGATTAACATGATATTAACAGTGAAGAGATGATATG 1637
Db 76341 ATTGATTAAGACATATTTCTTTAGAGACTATTTAGCTAGCGCTAAATATTAATCATGATC 76282
Oy 1638 TAAAGCTCATATATGAGGCAATGATGATGATGATGATGATGATGATGATGATGATGATG 1697
Db 76281 ACCGAGAAAAAGCTCTTAAGATTAAGATGATGATGATGATGATGATGATGATGATGATG 76222
Oy 1698 AAGTTGAGCTCAAGCCTTATCTAAGAAAAAGGATCTTCACTTCATCTCAGACGAG 1757
Db 76221 AGGACTCTAATTAAGCAGATG-----ACCAAGCTCTGACCACTGATATTAATCA 76168
Oy 1758 ATGTTAAGCAAAATCCAACTGAGATGATGAGAGAGATTTAATATGATGATGATGATG 1817
Db 76167 CATCAAAATTTAGTACAAAAAATTAAGTGTGAGAAATTTCAACAAACCAAAACCGAG 76108
Oy 1818 AAAAAGCAATTCAGCTGTTCTGATCTTCAATATATGATGATGATGATGATGATGATGATG 1877
Db 76107 AAAAAATTTGATCCCTGTTGATTAATGCTGCTCATGAGCATATGAGATGATGATGATG 76048
Oy 1878 ACGGTAATTTGATTTATTTCTCAATTAAGATCATTAATATTAATTTGCTGTTG 1937
Db 76047 ATGATCAATTTGATTTCTTCTCATATCATATTAATTTGCTATGAGCATGATGTTG 75988
Oy 1938 ---ATGATCAACATTAACAAAGCTCAAAATGCTTATCTGGAAGATTTGTTGCGACGA 1994
Db 75987 ACAAGGGGTGTTTATGGAACCAACAGAGCTTATTAACCAAACTCTTCTCAACAA 75928
Oy 1995 TTAAGTACTAGTGAACACCTGACGAACTTCAATATGATGATGATGATGATGATGATGATG 2054
Db 75927 TTAATATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 75871
Oy 2055 CCAAGTACATGTTTATGGAAGAAACCAAGTGAAGATCCAAATTAAGAACTTCAAG 2114
Db 75870 ACAGTATCATTAACAAAGGCTCAATTAAGCAATTAAGCAAAATTTATGCTCAGATG 75811
Oy 2115 CGGATGAAGACCAAGTGAAGAAACCTGCTGAGCCAGAGTCCCTCAAGTATGAGCTG 2174
Db 75810 AAGAACCTGTAAGATTCAGGGAAGTAACTCAACATATGTTTATGATGATTAATGAAG 75751
Oy 2175 AAAAAGTAGAAGCCCACTCAAGAGCAGAGAT 2209
Db 75750 GTTCAGACGAAGAAAGAACCAAGAAACCAAGAGAT 75716

RESULT 46
AX088372 2478 bp DNA Linear PAT 17-MAR-2001
LOCUS AX088372
DEFINITION Sequence 1 from Patent WO0114421.
ACCESSION AX088372
VERSION AX088372.1 GI:13397243
KEYWORDS
SOURCE
ORGANISM Streptococcus pyogenes
Streptococcus pyogenes
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 Koenig, S., Heinrichs, J., Johnson, L. S. and Adamou, J. E.
Homologs of a pneumococcal protein and fragments for vaccines
JOURNML Patent: WO 0114421-A 1 01-MAR-2001;
MEDLINE, INC. (US)
FEATURES
source
1. .2478
Location/Qualifiers
/organism="Streptococcus pyogenes"
/mol_type="genomic DNA"
/db_xref="taxon:1314"
BASE COUNT 846 a 486 c 518 g 628 t
ORIGIN
Query Match 9.0%; Score 214.6; DB 6; Length 2478;

```

Best Local Similarity 60.8%; Pred. No. 2.4e-35;
Matches 402; Conservative 0; Mismatches 229; Indels 30; Gaps 2;

QY	4	TTACAGAGTGGAGCTATCAAGCTAGAACGGTTAAGAAATTAATCGTGGTTCCATAT	63
DB	66	TTACCACTGGTGAACATCATATAGGTTACGACCAAGCAACAAATTAATGCTTATAT	125
QY	64	AGATGAAAAACAAGCGACGCAAAA-----ACGGAGATTGACTCTGATGAGGTAG	117
DB	126	TGATGTAGCAAGGTAAGCAAAAGCCCTTAAACAAACAAAGATGATCAATCAG	185
QY	118	CAAGGTGAAGGATATATGCTGAGCAATTCGTATCAAGATPAACAGCAAGGCTATGT	177
DB	186	TGCTGAGAAAGGATCTCTGCTGAACAGATCGAGTCAAAATTAATGACCAAGGCTATGT	245
QY	178	CACCTCAGATGGCGACCACTATCATTTATTAACAATGGTAGGTTCCCTTAGCGCTATCAT	237
DB	246	GACCTCAGATGGTGAACATTTATTTTAAATGGAAGTTCCCTTAGATGCGATTAAT	305
QY	238	CAGTGAAGATTAATCATGAAAGATCCAAACTATAAGCTAAAAAGATGAGATATTGTTAA	297
DB	306	TAGTGAAGAGTTGTTGATGACGATCTCTAATTAACCTTTTAAACATCAGAGCTTATCA	365
QY	288	TGAGGTCAAGGTTGATGATTTATCAAGTGAATGGAATACTATGTTTACCTTAAGA	357
DB	366	TGAATCTTAGACGGTTAGCTATTAAGTCAATGCGACATTAATGTTTACCTCAAGCC	425
QY	358	TGCTGCCCGACGGGATTAAGTCCGTTCAAAAGAGAAATCAATCGCAAAAACAGAGCA	417
DB	426	AGTGAAGCAAGCGCAAAACATTTGCAACCAACAAATTTGTCGACCAAGTGGCCAAAG	485
QY	418	TAGTCAACATCGTGAAG-----GTGAACTCCAAGAAAGCA	453
DB	486	AACCTAAGAGCTAAAGAAAAAGTTTACCTCAAGTGGCCCATCTTAAGTAAAGAGAGT	545
QY	454	TGCTGCTGTTGGCAGCTTTCGACAGAGCGCTATTAAGATGATGTTATATCTT	513
DB	546	TGCGGACAGTCAATGAAGCAAAAAGACAGAGCGCTATTAAGATGATGTTATTTT	605
QY	514	TAAATGTTTGAATCATGATGAGATGCTGATGTTATGCTTATGCTTCCATGAGATGA	573
DB	606	TAGTGAAGATCATGATGATGATTAAGAGATGTTATTAAGTACCTCATGATGATGA	665
QY	574	TTACATTAATCATCTTAAGATGATGATGATGATGATGATGATGATGATGATGATGATG	633
DB	666	CTATCATTAATTTCTTAAGATGATGATGATGATGATGATGATGATGATGATGATGATG	725
QY	634	C 634	
DB	726	C 726	

RESULT 47
AE006623/C 10825 bp DNA linear BCT 01-JUN-2001
LOCUS Streptococcus pyogenes M1 GAS strain SF370, section 152 of 167 of
DEFINITION the complete genome.
ACCESSION AE006623 AE004092
VERSION AE006623.1 GI:13623017
SOURCE Streptococcus pyogenes M1 GAS
STREPTOCOCCUS PYOGENES M1 GAS
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 (bases 1 to 10825)
Ferretti, J.J., McShan, W.M., Adji, D., Savic, D., Savic, G., Lyon, K.,
Primeaux, C., Seza, S.S., Surov, A.N., Kenon, S., Lai, H., Lin, S.,
Qian, Y., Jia, H.G., Najaf, F.Z., Ren, Q., Zhu, H., Song, L., White, J.,
Yuan, X., Clifton, S.W., Roe, B.A. and McLaughlin, R.E.
Complete genome sequence of an M1 strain of Streptococcus pyogenes
Proc. Natl. Acad. Sci. U.S.A. 98 (8), 4658-4663 (2001)
JOURNAL MEDLINE
PUBMED 11296296

REFERENCE 2 (bases 1 to 10825)
AUTHORS Ferretti, J.J., McShan, W.M., Adji, D., Savic, D., Savic, G., Lyon, K.,
Primeaux, C., Seza, S.S., Surov, A.N., Kenon, S., Lai, H., Lin, S.,
Qian, Y., Jia, H.G., Najaf, F.Z., Ren, Q., Zhu, H., Song, L., White, J.,
Yuan, X., Clifton, S.W., Roe, B.A. and McLaughlin, R.E.
TITLE Direct Submission
JOURNAL Submitted (10-APR-2001) Department of Microbiology and Immunology,
University of Oklahoma Health Sciences Center, 940 SL Young Blvd,
Oklahoma City, OK 73104, USA
FEATURES
source location/Qualifiers
1..10825
/organism="Streptococcus pyogenes M1 GAS"
/mol_type="genomic DNA"
/strain="SF370"
/serotype="M1"
/db_xref="taxon:160490"
complement (71..271)
/gene="SPY2005"
complement (71..271)
/note="Best Blastp hit = emb|CAA61195.1| (A132973)
hypotheetical protein SCF91.38c. [Streptomyces coelicolor
A3(2)]"
/codon_start=1
/transl_table=1
/product="hypotheetical protein"
/protein_id="AAK34687.1"
/db_xref="GI:13623018"
/translation="WSEKRSKIEQASGSGKEGAKULGDKLEAKFVEKTIARKG
ELADAKAEVAGAVDAVEKELK"
complement (440..2917)
/gene="SPY2006"
complement (440..2917)
/note="Best Blastp hit = gb|AAB58414.1| (AF000561) TTF-1
interacting peptide 21; TTF1; Transcription Termination
Factor 1 Interacting Peptide 21 [Homo sapiens]"
/codon_start=1
/transl_table=1
/product="hypotheetical protein"
/protein_id="AAK34688.1"
/db_xref="GI:13623019"
/translation="MKTYGYIGSVAAILLTHIGSYOLGKHHMGATKNOIAIYD
SKGAKAKTKMDOISAEIGISAEIOIVITDGGVTSQDHYHFNKGVPAVIAI
SEBLMTDPNRPKOSDYINELIDGYIKVANGYIVYIKPSKRNITKQOIAOVA
KGTFAKKGGLAQVAHLSKEVAAVNAEKRRGYITDGYIFSPDIIIDDADALVP
GHNHYVI.PKKDLSPSELAAQAVWSQKRGARSDYRTPAPAPGRKAPVPTVP
NPGQHPDNGSYHAPRPNDAQSKHQBDEKGTKEKELDQLRLDKRYVED
GLIPEPTOVIRKSNAGYVPHGDHYHII PRSOLSPLEMLADRYLAGOTEDDGSDD
SKPSDEYTHFLGRIKAYKGLDKGRPYSDAVPESKESHSYDKSGVTRAKGDPH
HYTGELAEQYELDEVANVAKGGADELAAALDQEGREKLPFTKVSRYTKDGK
VGYWPKKQKDYFARDDQLDLOIAFAOEMLDKRKARYDIVDTGLEPRADVSS
LPMHAGNATYDTSFVPIPHIDHIVVPSMLTRDQITIKYVMGPVPRPIMSKPG
HEESGAVIPNTPLDKRAGMNOIIEAEVQKALGRFATPGYIFEDPVLAKE
TFWKDGSFSPRADGSLRTITNSDKIAQMOAOELAKNAGADATDTTPKPKCO
ADKSNQGPSEBAKSEBEKSDPDIDSPVGLDRATLEDHINQAAKANDPKYLIF
QPEGYQFNKQELVITDITLQOINP"
complement (2930..3850)
/gene="1mb"
/note="synonym: SPY2007"
complement (2930..3850)
/gene="1mb"
/note="Best Blastp hit = emb|CAA5617.1| (X80397) orf2
[Streptococcus pyogenes]"
/codon_start=1
/transl_table=1
/product="putative laminin adhesion"
/protein_id="AAK34689.1"
/db_xref="GI:13623020"
/translation="MKGFFLMAMVSLMTAGCDKSNPKPTQKMSVTSFYPMYA
MTKEVSGDLNVRMIOGSGAIFHSFEPNDAAYADLPYHSHTEAMARLDLPLN
KSKVDPEASKPLTLDVKGLEMEVTEQIDPATLVPHTPVLVLEGEAVNAKKE
LGRIDPKKOSYTKNAKFAKKEAQLTREYVQKFKVSKTFVQHTAFSYLAKFGI

KOLGISISPEOEPSPROLKEIODFVEKENVKTIIPAEONVANPKIAHAIAKSTGAKVKT
 LSPLAAPSNGKTYLENLANLEVLQOLK"
 complement (4086. .5225)
 /gene="SPY2009"
 complement (4086. .5225)
 /gene="SPY2009"
 /note="Best Blastp hit = emb|CAA56616.1| (X80397) orf1
 [Streptococcus pyogenes]"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAK34690.1"
 /db_xref="GI:13623021"
 /translation="MRAENKNSHRSYIRKLKVGVTSAIAISFLGKVAVAVDGIPPI
 SLTQKTATSENHMHIDKGLIPLGISLEAKEFKEVESRSEKQKTYKOKIK
 TAPDKDLFTYHSEYTAIVVDLPASTSTTQPYEARPOEFOQASDSBMYGDSSTVT
 TDSPEETSSVPAAPALSEAPQAESEESVPAASEETSPTSPAPETPEEPAAP
 SPSESEPSVPAPEETPEEPEEPAPAPQAPAESESSVAATPSPTPASEETO
 TPNAVTKDSKSPSAAEKPAASLSVEQTVQOPTSKRSSDKKEQEOYSNRSLSRQ
 VRAHSGKYLSTGKQPLFIATWTLMSLFGSLLVTRQKETKK"
 complement (5322. .8867)
 /gene="scpa"
 /note="synonym: SPY2010"
 complement (5322. .8867)
 /gene="scpa"
 /note="Best Blastp hit = gb|AAB17762.1| (U56908) SCPB
 [Streptococcus agalactiae]"
 /codon_start=1
 /transl_table=11
 /product="C5A peptidase precursor"
 /protein_id="AAK34691.1"
 /db_xref="GI:13623022"
 /translation="MRKKOKLPDKLAIALMSTSLINQSDIKANTVEDTPATEOA
 VETPOPTVSESEBSSKETKTPQPDDEETIADNANLAQAAPAKTADPTATSKATI
 RDLNDSQVKTLOEKAGAGTVAIVADIGDKHAEARLTDKTKARIQSKEDLEKAK
 KEHGIITGEMWENVGLAYVHDYSKDGATVDQDEGTHVSGILSGNAPSTKPYRLEGA
 MPEAQLLMRVEIVGLADYARVNAQAIIDVNLGAKVIMNSFGNALAVANLPELGA
 KAFYPAKSGVSIVTSAGNDSFGGKTRPLADHDYGVGTGPAADSTLTPVASYPD
 KOLTEATKATADODDKEMPIVSTNRPEPNKAVDYAVANRMKDDPDVQKALILE
 RGDIDFKKINAKKAGAVGLIYDNOKGPRIELPNVDQMPAARISKDDLLKENP
 OKTTFNATPVLPASTGKLSRFSWGLTADGNLKPIIAPAGDILSVDEANRYAKL
 SGTSMAPLVAINGMLQKQYETQYPMVTPEERDLAKVIMSSATLAIYDEKNTYAKL
 PROQAGAVAKKASATMYVTDKNTISKVLNNVSQFVTVVHNSKSPQELPY
 OATQTDKVDGKLPALAKALYETSMOKITIPANSKOVTPIDVQSKOLAPMY
 GYPLGEPVRKODPTELMISIPYIGSPDGDNLSALEKPIYDSDGSSYYHEANSA
 KQJDDGGLGYTALKMNTALTITESNPTTIKAYKEGNIENEDIESSITETITAGT
 AKQDDSHYIHRHANGKPYVAISPNGGNDVVOFQETFLMANLVAELDKENV
 VMTSEVTEOVVKNVNDLASTIGSTRFERKTRMDKDKGVANGTYVRYVYPISS
 GAKQHTDFDIVNTTPEVATSAIFSTEDRRLTLASPKTSQPVYRRIAYTYMDE
 LPTTEYISPNEDGFTLPEBAETMGATVPLKMSDFTYVVEDMAGNITYVTKLLEG
 HSNKPEODGSDOAPDKKPEKTEPEDGSGOAKKPEKTEPEDGSGOCTPDKKPEKPEO
 DSGCOTPDKKPEKTEKDSGCGOTPEKTKQKOPSTTLEKRSKALATKASTKQDLPT
 TNDKOTNRLHLKLKLVMTTFPLGLVAHIFKTRTED"
 9301. .10506
 /gene="SPY2013"
 9301. .10506
 /gene="SPY2013"
 /note="Best Blastp hit = gb|AAC38767.1| (AF064540)
 putative transposase [Streptococcus pyogenes]"
 /codon_start=1
 /transl_table=11
 /product="transposase - IS1562"
 /protein_id="AAK34692.1"
 /db_xref="GI:13623023"
 /translation="MLHKEKPDYRNQGYFTLDQVDEHFLQVAVEAVIDFDEIYDL
 VEDTYSNGRPSLDPMVLVKIPLIQCFYGRSMQTKDIEVTAQVAYWFGILTLDDK
 VPHFTYGNYSRRPQEKGLIESIFTHVGLICITGILDPIDGTHIAAANNRK
 FINREVEKQAFMSOLEIENRDEKHKGLPAKLEPAVAKISITDDESGMFK
 GDHKEVPAYSAOVACDKGALAYSVENHDSQAPALAKLCEIPLKPOITIDSGY
 NVPSIAKTLIDKEITPULPYTRPRKDKLAPKELIYDEHDFCVICPHQQLITRTTT
 REGREYKSDPVICANCPILSVCTTSKNYQKVVTRHMKDYLAOPRNTTISAIRERA
 SPKWKIRLD"

ORIGIN
 Query Match 9.0%; Score 214.6; DB 1; Length 10825;
 Best Local Similarity 60.8%; Pred. No. 2.3e-35;
 Matches 402; Conservative 0; Mismatches 229; Indels 30; Gaps 2;
 Db 4 TTACGAGTTGGAGCTGATCAAGCTAGAACCGTTAGGAAATATGCTGTTCTTATAT 63
 2852 TTACCACTTGTAAGCATGATATGGGTTGACAAACAAAGACATCAATATGCTTATAT 2793
 Oy 64 AGATGAAAACACGACGACAAAA-----ACGAGATTGCTCCTGATGAGGTTAG 117
 2792 TGAATGATAGCAACAGTAAAGGCAAGCCCTTAAACAAACAAACGATGATCAATACG 2733
 Oy 118 CAACGTAAGGAATTCATATGCTGACAAATGTCATCAAGTAAACAGACCAAGGCTATGT 177
 2732 TGCTGAAGAAAGCATCTCTGCTGACAGATGATGTCAAAATTTCTGACCAAGGCTATGT 2673
 Db 178 CACTTCATGCGCAGCAGCATATCATTTATTAACATGTAAGGTTCTTATGACGCTATCAT 237
 2672 GACCTCATGTCATGACCATTTATCATTTTTCATGAGGAAAGTTCTTATGATGCGATAT 2613
 Oy 238 CAGTGAAATTAATCTCATGAAAGATCCAACTATTAAGCTAAAGATGAGATATTGTTAA 297
 2612 TAGTGAAGAGTTGTTGATGACGAGATCCTTAATACCGTTTAAACATCAGACGTTATCAA 2553
 Oy 298 TGAAGTCAAGGCTGATATGTTATCAAGGTAGAGAAATACATGTTTACCTTAAGGA 357
 2552 TGAATCTTGAACGCTTACGTTATTTAAATGCAATGAGCACTATTATGTTTACCTTAAGCC 2493
 Oy 358 TGCTGCCACCGCGATTAACGTCCTGACAAAAGAGAAATCAATGACAAACAAAGACA 417
 2492 AGGTAGCAAGGCGCAAAACATTTGACCAACAAACAAATGCTGAGCAAGTACGCAAAAG 2433
 Db 418 TAGTCAACATGCTGTAAG-----GTGAACCTCCAAGAAACGA 453
 2432 AACTTAAAGAAAGCTTAAAGAAAAAGTTAGCTCAAGTGGCCCATCTCGTAAAGAAAGAT 2373
 Oy 454 TGGTGCCTGTTGCTTGGCAGCTTGGCAAGGACGCTATACATGATGATGTTATATCTT 513
 2372 TGGCGAGCTCAATGACGAAAGAAAGACAAAGGCGTTATACATCAACATGCTTATATTTT 2313
 Db 514 TAACTCTTGATATCATATAGAGTACTGTGATGCTTATATCTTCTCATGAGAGATCA 573
 2212 TAGTCGACAGATATCATATGATGATTTAGAGAGATGCTTATTAATCACTCATGTAATCA 2253
 Oy 574 TTACCATTAATCTCTTAAGATGAGTATACGATGAGGAGTTGGCTGTCGAAAGCTT 633
 2252 CTATCATTAATATCTCTTAAAAAGATTGTCTCCAAAGTGAAGTACTGCTGCAACAGCTTA 2193
 Oy 634 C 634
 2192 C 2192
 Db
 RESULT 48
 AB073859
 LOCUS 2478 bp DNA linear BCT 03-NOV-2001
 DEFINITION Streptococcus pyogenes hepa gene for histidine triad protein of
 group A streptococci, complete cds.
 ACCESSION AB073859
 VERSION AB073859.1 GI:16610229
 KEYWORDS
 SOURCE Streptococcus pyogenes
 ORGANISM Streptococcus pyogenes
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.
 REFERENCE
 1 Terao, Y., Kawabata, S. and Hamada, S.
 Characterization of a novel histidine triad protein of group A
 streptococci
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2478)

BASE COUNT

2775 a 2207 c 2198 g 3645 t

AUTHORS Terao, Y., Kawabata, S. and Hamada, S.
TITLE Direct Submission
JOURNAL Submitted (01-NOV-2001) Yutaka Terao, Osaka University, Faculty of
Dentistry, Yamadaoka, 1-8, Suita, Osaka 565-0871, Japan
(E-mail: teraodent.osaka-u.ac.jp, Tel: 81-6-6879-2899,
Fax: 81-6-6878-4755)

FEATURES Location/Qualifiers

source
1. 2478
/organism="Streptococcus pyogenes"
/mol_type="genomic DNA"
/strain="SG1-9"
/serotype="M1"
/db_xref="taxon:1314"
1. 2478
/gene="hncpa"
1. 2478
/gene="hncpa"
/codon_start=1
/transl_table=1
/product="histidine triad protein of group A streptococci"
/protein_id="BAB71774.1"
/db_xref="GI:16610230"

CDS

BASE COUNT 846 a 488 c 518 g 626 t
ORIGIN
Query Match 8.9%; Score 213; DB 1; Length 2478;
Best Local Similarity 60.7%; Pred. No. 5.3e-35;
Matches 401; Conservative 0; Mismatches 230; Indels 30; Gaps 2;
4 TTACGAGTTGGGAGCTGATCAAGCTAAGACGTTAAAGAAATTCGTTCTTAT 63
66 TTACCACTGGTGAACATATATGGTTACGACCAAGACCAATCAATTCCTATAT 125
64 AGATGAAAACAAGCGACGCAAAA-----ACGAGATTTGACTCCTGATGAGTTAG 117
126 TGATCATATGCAAGGTAAGGCAAAAGCCCTAAACAAACAAAGATGATCAATCAG 185
118 CAAGGCTGAGGATTCATGCTGAGCAATTCCTCATCAAGATPAACAGCAAGCTATGT 177
186 TGCTGAAGAGGACATCTCTGCTGAACAGATCTAGCAAAATTAAGTCAAGAGCTATGT 245
178 CACTTCACATGGCGACCTATCATTTATTAAGGTAAGGTTAGTCTTATGAGCTATCAT 237
246 GACCTCACTGAGTGAACCATTTATTTTACATGGGAAAGTTCTTATGATGACATATAT 305
238 CAGTGAAGATTTACTCATGAAAGATCCAAACTATTAAGCTAAAGATGAGATATTGTTAA 297
306 TAGGCAAGAGTTGTTGATGAGGATCTTATTAACGTTTAAACAATCAGAGTTATCAA 365
298 TGAGGTCAAGGTTGATTTATTAAGGTAGATGGAATAATCTATGTTTACCTTAAAGA 357
366 TGAATCTTAGACGTTACGTTATTAAGTCAATGGCAACTTATTAATGTTTACCTCAAGCC 425
358 TGCTGCCCAAGCGGTAAGTCGCGTAACAAAGAGAAATCATTCGACCAAAAACAGAGAGA 417
426 AGGTGCAAGCGCAAAAACATTCGAACCAACAAATTCGCTGAGCAAGTACCAAGG 485
418 TAGTCAACATCGTGAAG-----GTGGAATCTCAAGAAACGA 453

DB 486 AACTAAGAGCTAAAGAAAAGTTAGTCAAGTGCCCATCTCAGTAAAGAGAGT 545
QY 454 TGTGCTCTTGGCTTGGACGCTTCCGACAGAGCGTATCTACAGTATGATGTTATATCT 513
DB 546 TCGGCACTCAATGAAAGAAAAGACAGGACGCTATCTACAGAGATGCTATATATTT 605
QY 514 TAATGCTCTGATATCATAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 573
DB 606 TAGTCCGACAGATATCATTTGATGATTTTGGAGATGATCTTATTTAGTACCTCATGTAATCA 665
QY 574 TTACCATTAATTCCTTAAGATGAGTTATGATGATGATGATGATGATGATGATGATGATGAT 633
DB 666 CTATCATATATATCTTAAGAAAGATTGTCTCAAGTAGAGTGTGCTGACCAAGCCCTA 725
QY 634 C 634
DB 726 C 726

RESULT 49
LOCUS AB010110 10618 bp DNA linear BCT 03-APR-2002
DEFINITION Streptococcus pyogenes strain MGAS8232, section 158 of 173 of the
complete genome.
ACCESSION AB010110 AE009949
VERSION AB010110.1 GI:19749144
KEYWORDS
SOURCE ORGANISM
Streptococcus pyogenes MGAS8232
Streptococcus pyogenes MGAS8232
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.

REFERENCE
1 (bases 1 to 10618)
Smoot, J.C., Barbican, K.D., Van Gompel, J.J., Smoot, L.M.,
Chaussee, M.S., Sylva, G.L., Sturdevant, D.E., Ricklefs, S.M.,
Porcella, S.F., Parkins, L.D., Beres, S.B., Campbell, D.S., Smith, T.M.,
Zhang, Q., Kapur, V., Daly, J.A., Veasy, L.C. and Musser, J.M.
Genome sequence and comparative microarray analysis of serotype M18
group A Streptococcus strains associated with acute rheumatic fever
outbreaks
Proc. Natl. Acad. Sci. U.S.A. 99 (7), 4668-4673 (2002)

JOURNAL MEDLINE 21927593
PUBMED 11917108
REFERENCE
2 (bases 1 to 10618)
Smoot, J.C., Barbican, K.D., Van Gompel, J.J., Smoot, L.M.,
Chaussee, M.S., Sylva, G.L., Sturdevant, D.E., Ricklefs, S.M.,
Porcella, S.F., Parkins, L.D., Beres, S.B., Campbell, D.S., Smith, T.M.,
Zhang, Q., Kapur, V., Daly, J.A., Veasy, L.C. and Musser, J.M.
Direct Submission
Submitted (31-JAN-2002) Laboratory of Human Bacterial
Pathogenesis/Rocky Mountain Laboratories/NIAID/NIH, 903 S. 4th St.,
Hamilton, MT 59840, USA

FEATURES Location/Qualifiers

source
1. 10618
/organism="Streptococcus pyogenes MGAS8232"
/mol_type="genomic DNA"
/strain="MGAS8232"
/db_xref="taxon:186103"
complement (95..259)
/gene="spym18..2063"
complement (95..259)
/gene="spym18..2063"
/note="best blastp match hypothetical"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAL98534.1"
/db_xref="GI:19749145"
/translation="MNIEVKKKISFVFPIINSFGATILLCFNITLSKEIILNFDI
LADYRMKLET"
216..917
/gene="SMEZ"
/note="synonym: spym18..2064"

216. .317
/gene="SMEZ"
/note="best blastp match gb|AAK34680.1| (AE006622)
mitogenic exotoxin Z [Streptococcus pyogenes M1 Gas]"
/codon start=1
/transl_table=11
/product="mitogenic exotoxin Z precursor"
/protein_id="AAL98535.1"
/db_xref="GI:19749146"
/translation="MKTKTLFSTFSIFAIIISRPVIGLEVDNNSLLRNISTIMYEY
SPTVIDFKSHLVTRKLDVARDAPFFINSEMEVYAADPFKDGKIAMFVSPFMNVL
SEGKVIAYVGGMTPEYKOTSIPKNI PVNLINGKDISVPIYIEISNKTIVAGSIDLK
VRKFLIAOHLSYSSGSSYSKSKLVFHTNDNSDLDFYGYGADKESIFVYVDKNS
ENIDITGHLIDIEIDS"
complement (1334. .1468)
/gene="spym18_2065"
complement (1334. .1468)
/gene="spym18_2065"
/note="best blastp match gb|AAK34681.1| (AE006622)
hypothetical protein [Streptococcus pyogenes M1 Gas]"
/codon start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL98536.1"
/db_xref="GI:19749147"
/translation="MTYFRLKELNWSLRNKGIVVEGNOQLQIAINTGNNSKSLF
"
1647. .325
/gene="dppA"
/note="best blastp match gb|AAK34682.1| (AE006622) surface
lipoprotein [Streptococcus pyogenes M1 Gas]"
/codon start=1
/transl_table=11
/product="surface lipoprotein"
/protein_id="AAL98537.1"
/db_xref="GI:19749148"
/translation="MSKYLKFSITTLFLGLIILVACCOQNPOTKEROKRPRODELV
VSMGKLPHEPDPKRDYGVHNEGNTHTSLTKRSPDLIKGLAKTVLSLDELTSF
DLHDFPESGPEPADVDVYKPTDMLKADGKAMDLTFIKVLEVNGKQNVNHLLEASH
TPTADLTETPIYPRKHNNDKYSNPISGSGPIVWVEYKAGDQAITVRNPVYHGKPYRR
KMTVALDLNLTALALBSGDVDMVYATPELASKVKVGTRLDIBSDNDRGSLSLPYVKRK
GVITPSKQVPGVNDVTSDPAIRKALTIGLRQKVLDTVLNGYGRPAYSIDKTPFMNVL
PTPTKPDNVPYAKQOLLTRKAGKEQADSRKKGDLDVAFDEYVPTNQLRANLIEVYAA
EBAKALGIIKILKASNMDEMAKTHSDSLVLAGGHHNQOPESSHPELACKGNTITIT
FNNNTLVTKVTLDKAMTSSDLDAKANEVYMLAOMDGKTASTIGLDLPNVLVLSNHTYIG
DKRINVGKGVSHSGHDSNLNTINIAETWDESTK"
3188. .4365
/gene="dppB"
/note="synonym: spym18_2067"
3388. .4365
/gene="dppB"
/note="best blastp match gb|AAK67218.1| (U78968)
transmembrane transport protein DppB [Streptococcus
pyogenes]"
/codon start=1
/transl_table=11
/product="transmembrane transport protein"
/protein_id="AAL98538.1"
/db_xref="GI:19749149"
/translation="MKRTIIIIWKIIRCTILFSGVSLTFEVLAKOSFVDPMVASVNVVY
DLSLPADYKALAHNYGLDPAALVOYFPTMLKQGHGTSLVVRQPSVDIISRAGAG
SFLMGSLGKAIISGLIGLITGSAFHOCKLDRIVRMSYVGLQISVPPWGLIILLIF
SVQLGWFPGLGISPTIGTISODITTLADRIKMLLPVFTISIIIGIANVTLHTTKMSVL
SSEVLVAAGKETQWQIFKHCLRNALNPVATTLHFSYFGLFGSGVLAIEVFSYFGI
GSTLFEAKRSDETPILLAI VMIGTLVFAGNLIADIINSINPOLRRKY"
4362. .5183
/gene="dppC"
/note="synonym: spym18_2068"
4362. .5183
/gene="dppC"

/note="best blastp match gb|AAK34684.1| (AE006622) transmembrane transport protein [Streptococcus pyogenes M1 GAS]"
 /codon_start=1
 /transl_table=1
 /product="transmembrane transport protein"
 /protein_id="AAL98539.1"
 /db_xref="GI:19749150"
 /translation="MILKRTVLMWLGIAISLILSLALNLFYRFRPLETNAALNRFAPSINLHGTGIDGDMFVRTIKGLYSLQVGLGALMGFLAVFVLNGLGSLIDKIIMLVDFLGMPHLFMIILSPVCKGAGGVIAAVVHPSLARLINREVDLKNKAPFOLSSMCKTPYVIIVRHIIILFVLSOIFIGILFPHVILHSAEMTFLGSLSAEOPSYGIISSEAKKHSISLGNMWLVIFPGYILVLVNAADTIGESLKKLIFYQTDHF"
 5195..5998
 /gene="qdpd"
 /note="synonym: spym18_2069"
 5195..5998
 /gene="qdpd"
 /note="best blastp match gb|AAK34685.1| (AE006622) ATPase protein [Streptococcus pyogenes M1 GAS]"
 /codon_start=1
 /transl_table=1
 /product="ATPase protein"
 /protein_id="AAL98540.1"
 /db_xref="GI:19749151"
 /translation="WTEITLISIKDLSTFTFOYGFELKFPSTPIQALNLYKKGELLAIIIGSGSKSLAAHIMDLIPKNAVTDGMIYRQSLSIKRIKQLRKEMTLLPQSVNLYLDSMKVYKHQVLRGISBNAKATQEGLFQOGLKESDGDYLPGLSGMRRVLFYTCISDTYSLIIADEPTPGHLPDALQVLDOLRFADKGSVIFITHDIVAASQIADRITIFKEKALTEETASPSFGSGEDQTEFARCLMRTLPGQDFLKGVTHDLRG"
 5982..6608
 /gene="qdpd"
 /note="synonym: spym18_2070"
 5982..6608
 /gene="qdpd"
 /note="best blastp match gb|AAK34686.1| (AE006622) ATPase protein [Streptococcus pyogenes M1 GAS]"
 /codon_start=1
 /transl_table=1
 /product="ATPase protein"
 /protein_id="AAL98541.1"
 /db_xref="GI:19749152"
 /translation="WTEIAKTLGFYHKKQDWLFREIDLEVAHQIIGIPGSGGKTSLSRVLAGLQKSGEVLVDGSHLPNKARFPVQLIQHPEQTMNLMPMKSLBEATYIFRSQDIRYAGIEQKMLNRRPSLSGGLQRFISVSLHPETKYLLADEMTLDSITQAPSVKSLIEIVEDRMGLIVISHPEFMALEKLCDACMYTEENRTOLFKAH"
 complement (6690..6890)
 /gene="spym18_2071"
 complement (6650..6850)
 /gene="spym18_2071"
 /note="best blastp match gb|AAK34687.1| (AE006623) hypothetical protein [Streptococcus pyogenes M1 GAS]"
 /codon_start=1
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="AAL98542.1"
 /db_xref="GI:19749153"
 /translation="MSEKELKSIQASGGLKEGAKLTGDKLEAKGFYEKTIYAKGKELADAKAVEGADVAVKEKXK"
 complement (7059..9530)
 /gene="spym18_2072"
 complement (7059..9530)
 /gene="spym18_2072"
 /note="best blastp match gb|AAK34688.1| (AE006623) hypothetical protein [Streptococcus pyogenes M1 GAS]"
 /codon_start=1
 /transl_table=1
 /product="hypothetical protein"
 /protein_id="AAL98543.1"
 /db_xref="GI:19749154"
 /translation="MKKTYGIGSVAAILATHTIGSYOLGKHMGVYKQNOIAYIDD SKGAKAKTKTKMDQISABEGISAEQIVKTIQDGYVLSGHDHYHYNGVNPDAIIDD SKGAKAKTKTKMDQISABEGISAEQIVKTIQDGYVLSGHDHYHYNGVNPDAIIDD SEELLMDPNRIFQSDVINELDGVIKQNGNYYVVLKPSKSKNLRITKQOIAEYVA

KGTKEAKKGLAQAVALHLSKEVAANEAQRQRYTTDDGYIFSPDIIIDLDGAVLVP
HGNHYHYPKDLSPBELAAQAQYWSQKQDEFPKPTKEILLDQHLRLDKYRVVEEDGL
RQGHQPDNGYHPAPRRPNDASQNKQKQDEFPKPTKEILLDQHLRLDKYRVVEEDGL

Query Match 8.8%; Score 213; DB 1; Length 10618;
Best Local Similarity 60.7%; Pred. No. 5.1e-35;
Matches 401; Conservative 0; Mismatches 230; Indels 30; Gaps 2;

```

QY 4 TTACGAGTTGGGACCTGATCAAGCTAGAACGGTTAAGGAAATATCGTGTTCCTATAT 63
DB 9465 TTACCGAGCTTGGTAGACATCATATGAGTCCAGTAAACAAAGCAATCAGATTCCTATAT 9406
QY 64 AGATGAAAAACAAGCGACGCAAAA-----ACGGAATTGACTCTCGATGAGGTTAG 117
DB 9405 TGATGATAGCAAAAGTAGGCAAAAAGCCCTTAAACAAACAAAGATGATCAATCG 9346
QY 118 CAAGCTGAGGAATATATGCTGAGCAATTCGTATCAAGATTAACGACCAAGCTATGT 177
DB 9345 TCGTGAAGAGGATCTCTGCTGAACAGATCGTAGCAAAATTAAGTACCAAGGTTATGT 9286
QY 178 CACTTCACATGGGACACCATCATATTAACAATGTAAGTTCCCTTAAGGCTATCAT 237
DB 9285 GACCTTACACGGTGACCATTAATTTTACAAATGGGAAAGTTCTTAATGATGCGATTA 9226
QY 238 CAGTGAAGATTACTCATGAAAGATCCAAACTAATAGCTAAAGATGAGATATGTTAA 297
DB 9225 TAGTGAAGATTGTTATGATGCGATCTTAATTCGTTTAAACATCAGAGCTTACA 9166
QY 298 TGAAGTCAAGGTTGATATGTTATCAAGTAGATGGAATAATCTATGTTTACCTTAAGA 357
DB 9165 TGAATCTTAGACGGTTAGCTTAATTAAGTCAATGCAACTTAATGTTTACCTCAAGCC 9106
QY 358 TCGTGCCGACGGGATTAAGTCGCTACAAAGAGAAATCAATCCGCAAAAACAAAGCA 417
DB 9105 AGCTACCAAGCGCAAAAACATTTGAACAAACAAATTCGACAGCAAGTACCAAGG 9046
QY 418 TAGTCAACTCGTAG-----GTGAATCTCAAGAAACGA 453
DB 9045 AACTAAGAGCTGAAGAAAGAGTTAGTCAAGTGGCCCATCTCAGTAAAGAAAGAT 8986
QY 454 TGGTGTGTTGCTTGGCAGCTTCCGACAGGCTTACTACAGATGATGTTATCTT 513
DB 8985 TCGCGAGAGTCAATGAAGCAAAAAGACAGACCTTACTACAGAGATGTTATCTT 8926
QY 514 TAATGTTCTGATCATGAGGATCTGCTGATGCTTATGCTTCCCTCAGAGATCA 573
DB 8925 TAGTCCGACAGATATGATGATTTAGAGACGCTTATTAATTAATCTCAGTATCA 8866
QY 574 TTACCATTAATCTCTAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 633
DB 8865 CTATCATTAATTTCTTAAGAAAGATTTGCTCCAAAGTGAAGTGTGCTGACAGCTTA 8806
QY 634 C 634
DB 8805 C 8805

```

RESULT 50
SAG290952 1910 bp DNA linear BCT 07-JUL-2002
LOCUS Streptococcus agalactiae lmb gene for laminin-binding surface
DEFINITION protein, partial ORFy and partial insertion sequence IS1548.
ACCESSION AJ290952
VERSION AJ290952.1 GI:13548638
KEYWORDS adhesin; insertion sequence IS1548; laminin-binding surface
protein; lmb gene; orfy.
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.

REFERENCE
AUTHORS Granlund,M., Michel,F. and Norgren,M.
TITLE Mutually exclusive distribution of IS1548 and GBS11, an active

group II intron identified in human isolates of group B
streptococci
J. Bacteriol. 183 (8), 2560-2569 (2001)
JOURNAL
MEDLINE 21172873
PUBMED 11274116
REFERENCE 2 (bases 1 to 1910)
AUTHORS Granlund,M.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-2000) Granlund M., Dept. of Clinical
Bacteriology, Clinical Microbiology, Umea University, S-901 85
Umea, SWEDEN

FEATURES

```

source
1..1910
/organism="Streptococcus agalactiae"
/mol_type="genomic DNA"
/isolate="531"
/db_xref="taxon:1311"
/country="Sweden:Umea"
/note="serotype III"
1..30
/note="serotype III"
insertion_region
21..30
/insertion_seq="IS1548"
repeat_region
31..39
/rpt_type=DIRECT
misc_feature
31..39
/note="spacer region"
gene
40..1031
/gene="lmb"
40..45
/gene="lmb"
63..68
/gene="lmb"
111..1031
/gene="lmb"
111..1031
/function="putative adhesin"
CDS
/codon_start=1
/transl_table=1
/product="laminin-binding surface protein"
/protein_id="CAC35984.1"
/db_xref="GI:13548639"
/db_xref="SPTREMBL:Q9AE22"
/transl_table=1
/product="MKKGFLMAMVSLVMIAGCDKSNPKPTGMSVYTFYPMYA  
MKVEGSLDPMVROSAGLHSEFSPVNDVAIYDADLFVYHSHLEAMADLDNL  
KSKYDVEASKEPLTLDRVKGLEHMEVTOGIDPATLDPHTMTDVLAGEEAVNAKE  
LGRDLPKXKSTYTKAKAFKXEBOLTEYTKPKYVSKTPTVOTHTAFSTYAKRGL  
KOLGSGISPEDESPROKETOYVKEVNTTFAEDNVNPKIAHAIKSTGAVKT  
LSPLEAASGNKTYLENLRLNLEVLVYQOLK"
1044..>1910
/note="ORFy"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="CAC35985.1"
/db_xref="GI:13548640"
/db_xref="SPTREMBL:Q9AE21"
/transl_table=1
/product="MKKYGYIGSYAAILLTHIGSYOLOGHMGHLATKNOIAYID  
SKGAKAPKTKNDQISAEBGISAEQIVKATIDGIVTSHGDHFTNGKVPYALI  
SEELMTDPMYRFKOSDIVINELTDGVYIKVGNVYVYLPKSGKRNIRKQOIAOVA  
KGTKEAKKGLAQAVALHLSKEVAANEAQRQRYTTDDGYIFSPDIIIDLDGAVLVP  
HGNHYHYPKDLSPBELAAQAQYWSQKQDEFPKPTKEILLDQHLRLDKYRVVEEDGL  
RQGHQPDNGYHP"

```

CDS

BASE COUNT 651 a 366 c 404 g 489 t

Query Match 8.8%; Score 211.4; DB 1; Length 1910;
Best Local Similarity 60.5%; Pred. No. 1.2e-34;
Matches 400; Conservative 0; Mismatches 231; Indels 30; Gaps 2;

```

QY 4 TTACGAGTTGGGACCTGATCAAGCTAGAACGGTTAAGGAAATATCGTGTTCCTATAT 63
DB 1109 TTACCGAGCTTGGTAGACATCATATGAGTCCAGTAAACAAAGCAATCAGATTCCTATAT 1168
QY 64 AGATGAAAAACAAGCGACGCAAAA-----ACGGAATTGACTCTCGATGAGGTTAG 117

```


C 82 45.4 1.9 58985 25 ABZ59738 Human secreted pro
C 83 45.4 1.9 143601 25 ABZ22654 Human epididymal gr
84 45.2 1.9 1303 22 AAV29882 C albicans apoptos
85 45.2 1.9 2277 19 AAV13834 Homo sapiens ambly
86 45.2 1.9 2277 19 AAV05370 Homo sapiens ambly
87 44.6 1.9 423 22 ABA48060 Human breast cell
88 44.6 1.9 423 22 ABA65943 Human foetal liver
89 44.6 1.9 423 22 ABA33035 Probe #11501 for g
90 44.6 1.9 423 22 ABA14364 Human brain expres
91 44.6 1.9 423 22 AAK40093 Human bone marrow
92 44.6 1.9 423 22 AAI20873 Probe #10806 for g
93 44.6 1.9 423 22 AAI46111 Probe #14797 used c
94 44.6 1.9 423 22 AAI06583 Probe #6574 used c
95 44.6 1.9 423 23 ABS39677 Human liver single
96 44.6 1.9 423 24 ABS14168 Human genome-deriv
97 44.6 1.9 487 22 ABA42932 Human breast cell
98 44.6 1.9 487 22 ABA53351 Human foetal liver
99 44.6 1.9 487 22 ABA21131 Probe #1597 for ge
100 44.6 1.9 487 22 AAK01618 Human brain expres

ALIGNMENTS

RESULT 1
AAV27351
ID AAV27351 standard; DNA; 2389 BP.

XX AAV27351;

DT 02-OCT-1998 (first entry)

XX Streptococcus pneumoniae SP0036 nucleotide.

KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KM detection; pneumonia; Otitis media; meningitis; ss.

OS Streptococcus pneumoniae.

XX Key Location/Qualifiers

FT CDS 2..2389

FT /*tag= a

FT /product= "SP0036"

FT /transl_except= (pos:1367..1369,aa:Xaa)

FT /note= "No stop codon given; Xaa is unspecified"

XX MO9818930-A2.

XX 07-MAY-1998.

XX 30-OCT-1997; 97WO-US19422.

XX 31-OCT-1996; 96US-0029960.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Choi GH, Hromockyj A, Johnson LS, Kunsch CA;

XX WPI; 1998-272224/24.

XX 3-PSDB; AAM55090.

XX Nucleic acid encoding antigenic peptide(s) from Streptococcus

XX pneumoniae - or their epitope-containing fragments, useful in

XX protective or therapeutic vaccines, and for diagnosis

XX Claim 1; Page 59; 118pp; English.

XX The present sequence encodes a protein from Streptococcus pneumoniae.

XX The nucleic acid sequence encoding the Streptococcus pneumoniae protein

XX can be useful in vaccines for inducing protective antibodies against

XX Streptococcus pneumoniae, for treatment or prevention of infection e.g.

XX pneumonia, otitis media or meningitis. Probes based on the nucleic acid

XX are used to detect Streptococcus infection (by usual hybridisation or

CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
CC (especially 10-300) mu g/ml per dose.

XX Sequence 2389 BP; 830 A; 461 C; 486 G; 611 T; 1 other;

Query Match 100.0%; Score 2388; DB 19; Length 2389;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGAGCTGTAATCAAGCTAGAACCGTTAAGAAATATGCTTTCTTA 60
Db 1 TTCTTACGAGTTGGAGCTGTAATCAAGCTAGAACCGTTAAGAAATATGCTTTCTTA 60
QY 61 TATAGTGAAGAAACACGAGCGCAAAACGAGAAATTTGACTCTGATGAGTTAGCA 120
Db 61 TATAGTGAAGAAACACGAGCGCAAAACGAGAAATTTGACTCTGATGAGTTAGCA 120
QY 121 GCGTGAAGAAATCAATGCTGAGCAAAATGCTATCAAGATTAACAGCAAGCTATGTCAC 180
Db 121 GCGTGAAGAAATCAATGCTGAGCAAAATGCTATCAAGATTAACAGCAAGCTATGTCAC 180
QY 181 TTCACATGGGACCACTATCATTTTCAATGTGTAAGTTCTTATGACCTATCATCAG 240
Db 181 TTCACATGGGACCACTATCATTTTCAATGTGTAAGTTCTTATGACCTATCATCAG 240
QY 241 TGAAGATTTACTCATGAAAGTCCAAACTTAAGCTAAAGATGAGATATTTGTAATGA 300
Db 241 TGAAGATTTACTCATGAAAGTCCAAACTTAAGCTAAAGATGAGATATTTGTAATGA 300
QY 301 GGTCAAGGTGATATGTTTCAAGGTAGATGAAATTAATCTTATGCTTAAGATGC 360
Db 301 GGTCAAGGTGATATGTTTCAAGGTAGATGAAATTAATCTTATGCTTAAGATGC 360
QY 361 TGCCACGCGGATTAAGTCCGCTACAAAGAGAAATCAATCGAAGAAACAAAGGCAATAG 420
Db 361 TGCCACGCGGATTAAGTCCGCTACAAAGAGAAATCAATCGAAGAAACAAAGGCAATAG 420
QY 421 TCAACATCGTGAAGGTGAATCCCAAGAAACGATGCTGTTGCTTGGACCTTGCA 480
Db 421 TCAACATCGTGAAGGTGAATCCCAAGAAACGATGCTGTTGCTTGGACCTTGCA 480
QY 481 AGAAGCTATATCAAGATGATGTTATCTTAAATGCTTCTGATATCATAGAGATAC 540
Db 481 AGAAGCTATATCAAGATGATGTTATCTTAAATGCTTCTGATATCATAGAGATAC 540
QY 541 TGGTGAATGCTTATATGCTTCTCAATGAGATCATTTACATTTACATTTGAGATT 600
Db 541 TGGTGAATGCTTATATGCTTCTCAATGAGATCATTTACATTTACATTTGAGATT 600
QY 601 ATCAGTAGAGAGTTGGCTGCTGAGAAAGCTTCTTATGCTGAGAGAAATGCTCAAA 660
Db 601 ATCAGTAGAGAGTTGGCTGCTGAGAAAGCTTCTTATGCTGAGAGAAATGCTCAAA 660
QY 661 TTCAAGAACTTATGCGCCGACAAATAGCATTAACCTTCAAGAACTTGGGATACCTTC 720
Db 661 TTCAAGAACTTATGCGCCGACAAATAGCATTAACCTTCAAGAACTTGGGATACCTTC 720
QY 721 TGTAGCAATTCAGAACTCAAAATATCAACAGCAACAACTAATCACTCA 780
Db 721 TGTAGCAATTCAGAACTCAAAATATCAACAGCAACAACTAATCACTCA 780
QY 781 AGCAAGTCAAGTAATATGATGCTTCTTGAACAGCTTACAACTGCTTTGAG 840
Db 781 AGCAAGTCAAGTAATATGATGCTTCTTGAACAGCTTACAACTGCTTTGAG 840
QY 841 TCAACGATGTAGAAATGATGCTTCTTGAACAGCTTACAACTGCTTTGAG 900
Db 841 TCAACGATGTAGAAATGATGCTTCTTGAACAGCTTACAACTGCTTTGAG 900

Db 1921 TAAATTGCTTGGTTGATGATCACAATACAAAGCTCCAAATGGCTATACCTTGGAGA 1980
QY 1981 TTTGTTGGGACGATTAAGTACTAGTAGAACAACCTGAGCAAGCTCCACATTTCAATGA 2040
Db 1981 TTTGTTGGGACGATTAAGTACTAGTAGAACAACCTGAGCAAGCTCCACATTTCAATGA 2040
QY 2041 TGGATGGGCAATGCGCAGTAGCATGTGTTAGGCAAGAAAGACACAGTGAAGATCCAA 2100
Db 2041 TGGATGGGCAATGCGCAGTAGCATGTGTTAGGCAAGAAAGACACAGTGAAGATCCAA 2100
QY 2101 TAAAGACTTCAAAAGCGGATGAGAGACCACTGCTGAGCCAGAAATGCC 2160
Db 2101 TAAAGACTTCAAAAGCGGATGAGAGACCACTGCTGAGCCAGAAATGCC 2160
QY 2161 TCAAGTAGAGACTGAAAGAGTAGAGCCCACTCAAGAGAGAGAGAGTTTGTTCGGA 2220
Db 2161 TCAAGTAGAGACTGAAAGAGTAGAGCCCACTCAAGAGAGAGAGAGTTTGTTCGGA 2220
QY 2221 AGTAAAGGATTTAGTCTGAAAGCCAAATGCAAGAACTTGTGTTTACGAATTA 2280
Db 2221 AGTAAAGGATTTAGTCTGAAAGCCAAATGCAAGAACTTGTGTTTACGAATTA 2280
QY 2281 TTTGACTCTTCAATTTATGATATACAAATGATATGCGCAAGAGCAAAATTAATTCGC 2340
Db 2281 TTTGACTCTTCAATTTATGATATACAAATGATATGCGCAAGAGCAAAATTAATTCGC 2340
QY 2341 GTTGTAAAGAGAGTATCTTCATCTGTAGTAGAGAAATAATTAAC 2389
Db 2341 GTTGTAAAGAGAGTATCTTCATCTGTAGTAGAGAAATAATTAAC 2389

RESULT 3
ABX06886
ID ABX06886 standard; DNA; 2406 BP.

XX ABX06886;

XX 11-FEB-2003 (first entry)

XX S. pneumoniae type 4 strain coding region #1174.

XX Gene; d; bacterial meningitis; pneumonia; sepsis; otitis media;
XX ear infection; antiinflammatory; antibacterial; immunostimulant;
XX auditory; respiratory; gene therapy; vaccine.

XX Streptococcus pneumoniae type 4 strain.

XX WO200277021-A2.

XX 03-OCT-2002.

XX 27-MAR-2002; 2002WO-1B02163.

XX 27-MAR-2001; 2001GB-0007658.

XX (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

XX Masignani V, Tettelein H, Fraser C;

XX WPI; 2003-040579/03.

XX P-PSDB; ABU01598.

XX New proteins and nucleic acid molecules from Streptococcus pneumoniae,
XX useful as medicaments for treating or preventing a disease or infection
XX due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
XX or ear infection -
XX Claim 6; SEQ ID No 2347; 56pp; English.

XX The invention relates to a protein comprising or having at least 50%
XX identity to any of the 2469 amino acid sequences, identified in the
XX specification (available on a computer readable format), or its fragment,

CC expressed from 2469 of 2489 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC AB556454. Also included are an antibody which binds one of the
CC proteins, treating a patient by administering the protein, DNA or
CC antibody (in a composition), a kit comprising first and second primers,
CC which are the nucleic acid cited above or fragments between nucleotides
CC 8-100 of a sequence not defined in the specification, for amplifying a
CC target sequence contained within a Streptococcus nucleic acid sequence,
CC where the first primer is substantially complementary to the target
CC sequence and the second primer is substantially complementary to the
CC complement of the target sequence, and where the parts of the primers
CC having substantial complementarity define the termini of the target
CC sequence to be amplified, assay comprising contacting a test compound
CC with the protein, and determining whether the test compound binds to the
CC protein and a Streptococcus pneumoniae bacterium, where one or more
CC genes encoding the proteins has been rendered inactive. The proteins,
CC nucleic acid molecules, antibody and compositions are useful as
CC medicaments for treating or preventing a disease or infection due to
CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
CC sepsis, otitis media or ear infection. They are also useful in developing
CC vaccines, diagnostic and antibiotics. The methods are useful for
CC identifying immunodominant proteins. The present sequence is one of
CC the 2489 identified coding region from the genomic sequence.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.

XX .Sequence 2406 BP; 834 A; 461 C; 490 G; 621 T; 0 other;

Query Match 100.0%; Score 2388; DB 25; Length 2406;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGAGCTGTATCAAGCTTGAACGGTTAAAGAAATATCGTGTCTTA 60
Db 18 TTCTTACGAGTTGGAGCTGTATCAAGCTTGAACGGTTAAAGAAATATCGTGTCTTA 77
QY 61 TTATAGTGAAGAAACGAGCGCAAAACGAGAAATTTGACTCCGTAGAGTTAGCA 120
Db 78 TTATAGTGAAGAAACGAGCGCAAAACGAGAAATTTGACTCCGTAGAGTTAGCA 137
QY 121 GCGTGAAGGAATCAATGTCTGAGCAAAATGCTCATCAAGATTAAGACCAAGCTATGTC 180
Db 138 GCGTGAAGGAATCAATGTCTGAGCAAAATGCTCATCAAGATTAAGACCAAGCTATGTC 197
QY 181 TTCAATGCGGACCACTATCATTTATTAACATGTAAGGTTCTTATGACGCTATCATCG 240
Db 198 TTCAATGCGGACCACTATCATTTATTAACATGTAAGGTTCTTATGACGCTATCATCG 257
QY 241 TGAAGATTTACTCATGAAGAAATCCAAACTATTAAGCTTAAAGATGAGATTTGTAATGA 300
Db 258 TGAAGATTTACTCATGAAGAAATCCAAACTATTAAGCTTAAAGATGAGATTTGTAATGA 317
QY 301 GGTCAAGGTTGATATGTTATCAAGTATGAGAAATATCTATGTTTACCTTAAGATGC 360
Db 318 GGTCAAGGTTGATATGTTATCAAGTATGAGAAATATCTATGTTTACCTTAAGATGC 377
QY 361 TGCCCAAGCGGATTAAGCTTCGTAACAAAGAGAAATCAATGCAACAAAGAGATAG 420
Db 378 TGCCCAAGCGGATTAAGCTTCGTAACAAAGAGAAATCAATGCAACAAAGAGATAG 437
QY 421 TCAACATCGTGAAGGTGAATCCCAAGAAACGATGCTGTGCTTGGCAGCTTGGCA 480
Db 438 TCAACATCGTGAAGGTGAATCCCAAGAAACGATGCTGTGCTTGGCAGCTTGGCA 497
QY 481 AGGACGCTATATCAAGATGATGTTATATCTTTATGCTTGTGATATCATAGAGATAC 540
Db 498 AGGACGCTATATCAAGATGATGTTATATCTTTATGCTTGTGATATCATAGAGATAC 557
QY 541 TGGTATGCTTATATGTTCTCATGAGATCATTAACATTAATCTTCAAGATGAGTT 600
Db 558 TGGTATGCTTATATGTTCTCATGAGATCATTAACATTAATCTTCAAGATGAGTT 617

Oy 601 ATCAGCTAGCAGTGTGGCTGCTGAGAAAGCCTTCTATCTGGTCGAGGAAATCTGTCAA 660
 Db 618 ATCAGCTAGCAGTGTGGCTGCTGAGAAAGCCTTCTATCTGGTCGAGGAAATCTGTCAA 677
 Oy 661 TTCAAGAACTATGCGCCGACAAATATGCGATTAACATTCAGAAACAACTGGGATCCTTC 720
 Db 678 TTCAAGAACTATGCGCCGACAAATATGCGATTAACATTCAGAAACAACTGGGATCCTTC 737
 Oy 721 TGTAGCAATCCAGAACTACAAATATCTAACACAAAGCAACAGCAACTAACAGTCA 780
 Db 738 TGTAGCAATCCAGAACTACAAATATCTAACACAAAGCAACAGCAACTAACAGTCA 797
 Oy 781 AGCAAGTCAAGATTAATGATCATTTAGTCTGTGAACAGCTCTCAAACTGCCCTTGTAG 840
 Db 798 AGCAAGTCAAGATTAATGATCATTTAGTCTGTGAACAGCTCTCAAACTGCCCTTGTAG 857
 Oy 841 TCAACGACATGTAGATCTGATGGCTTGTCTTTGATCCAGACAAATCAGAGTCCGAA 900
 Db 858 TCAACGACATGTAGATCTGATGGCTTGTCTTTGATCCAGACAAATCAGAGTCCGAA 917
 Oy 901 AGCTAGAGTGTGCGAGTCCACAGGAGATCATTAACATTCCTTACTCTCAAT 960
 Db 918 AGCTAGAGTGTGCGAGTCCACAGGAGATCATTAACATTCCTTACTCTCAAT 977
 Oy 961 GTCTGAATTTGAGAAAGCAATCGCTCGATTAATTCCTGCTTATCGTTCAAACTATG 1020
 Db 978 GTCTGAATTTGAGAAAGCAATCGCTCGATTAATTCCTGCTTATCGTTCAAACTATG 1037
 Oy 1021 GGATACAGATTCAGAGCCAGAAACAAAGTCCACAACTCCGAACTAGTCCAG 1080
 Db 1038 GGATACAGATTCAGAGCCAGAAACAAAGTCCACAACTCCGAACTAGTCCAG 1097
 Oy 1081 CCGGCAACTGCAACCAATCTTAAATATGACTCAAACTCTTCTTTGGTATGCTAGCTG 1140
 Db 1098 CCGGCAACTGCAACCAATCTTAAATATGACTCAAACTCTTCTTTGGTATGCTAGCTG 1157
 Oy 1141 ACGAAAGTTGGGAGAGATATGTTATCGAAGAAAGGCACTCTGTTATGCTTTC 1200
 Db 1158 ACGAAAGTTGGGAGAGATATGTTATCGAAGAAAGGCACTCTGTTATGCTTTC 1217
 Oy 1201 GAAAGATTTACATCTGAAATCTTAAATCTTGAAGCAAGTTATCAAAACAGAGAG 1260
 Db 1218 GAAAGATTTACATCTGAAATCTTAAATCTTGAAGCAAGTTATCAAAACAGAGAG 1277
 Oy 1261 TGTTCACACACTTAACTGCTAAAGAAATGTTGCTCTCTGTCGCAAGATTTTA 1320
 Db 1278 TGTTCACACACTTAACTGCTAAAGAAATGTTGCTCTCTGTCGCAAGATTTTA 1337
 Oy 1321 TGATTAAGCATATATCTGTTAACTGAGGCTCATTAAGCCTTGTGNAATTAAGGTCG 1380
 Db 1338 TGATTAAGCATATATCTGTTAACTGAGGCTCATTAAGCCTTGTGNAATTAAGGTCG 1397
 Oy 1381 TAAATTCGATTTCCAGGCTTAGACAAATTAATTAGAACGCTTGAATGATGACATTA 1440
 Db 1398 TAAATTCGATTTCCAGGCTTAGACAAATTAATTAGAACGCTTGAATGATGACATTA 1457
 Oy 1441 TAAAGAAAAATTTGATGATGATTTATTTGCAATTCCTAGACCAATTAACCTCCAGCG 1500
 Db 1458 TAAAGAAAAATTTGATGATGATTTATTTGCAATTCCTAGACCAATTAACCTCCAGCG 1517
 Oy 1501 ACTTGGCAAAACCAATTTCAAAATGAGTATACGAAAGCAAGTGTGATTTGCAATT 1560
 Db 1518 ACTTGGCAAAACCAATTTCAAAATGAGTATACGAAAGCAAGTGTGATTTGCAATT 1577
 Oy 1561 AGCTGATAGTATACAAAGTGCAGATGTTACATTTTGTATGAACATGATATATCAGTGA 1620
 Db 1578 AGCTGATAGTATACAAAGTGCAGATGTTACATTTTGTATGAACATGATATATCAGTGA 1637
 Oy 1621 TGAAGGAGATGATATGTAAGGCTCATATGCGCCATATGTCACCTGATTTGAAAAAGATAG 1680
 Db 1638 TGAAGGAGATGATATGTAAGGCTCATATGCGCCATATGTCACCTGATTTGAAAAAGATAG 1697

Oy 1681 CCTTTCGATTAAGAAAAAGTTGACGCTCAAGCCTATATTAAGAAAAAGTATCTTACC 1740
 Db 1698 CCTTTCGATTAAGAAAAAGTTGACGCTCAAGCCTATATTAAGAAAAAGTATCTTACC 1757
 Oy 1741 TCCATCTCCAGACGCAATGTTAAAGCAATCCAACTGAGATGTCAGACGATATTTA 1800
 Db 1758 TCCATCTCCAGACGCAATGTTAAAGCAATCCAACTGAGATGTCAGACGATATTTA 1817
 Oy 1801 CAATCGTGTGAAGGGGAAAAAGAAATTCACCTGCTGAGCTTCATATATGTTGAGCA 1860
 Db 1818 CAATCGTGTGAAGGGGAAAAAGAAATTCACCTGCTGAGCTTCATATATGTTGAGCA 1877
 Oy 1861 TACAGTTAGGTTAAAAACGGTAATTTGATTAATTCCTATATAGATCATTAACATATAT 1920
 Db 1878 TACAGTTAGGTTAAAAACGGTAATTTGATTAATTCCTATATAGATCATTAACATATAT 1937
 Oy 1921 TAAATTTGCTTGTGTTATGATATCACATACAAAGCTCCAAATGGCTATATCTTGAAGA 1980
 Db 1938 TAAATTTGCTTGTGTTATGATATCACATACAAAGCTCCAAATGGCTATATCTTGAAGA 1997
 Oy 1981 TTTGTTGGCAGCATTAAGTACTAGTAAAGCAACCTGAGAAAGTCCACATTTCTAATGA 2040
 Db 1998 TTTGTTGGCAGCATTAAGTACTAGTAAAGCAACCTGAGAAAGTCCACATTTCTAATGA 2057
 Oy 2041 TGGATGGGCAATGECAGTGAAGCATGTTAGGCAAGAAAGCCACAGTGAAGATCCAAA 2100
 Db 2058 TGGATGGGCAATGECAGTGAAGCATGTTAGGCAAGAAAGCCACAGTGAAGATCCAAA 2117
 Oy 2101 TAAGACTTCAAACCGGATGAGAGCCAGTGAAGAAACCTCTGTCAGCCAGAAATCCC 2160
 Db 2118 TAAGACTTCAAACCGGATGAGAGCCAGTGAAGAAACCTCTGTCAGCCAGAAATCCC 2177
 Oy 2161 TCAAGTGAAGCTGAAAAAGTGAAGGCCCACTCAAGAAAGCAGAAAGTTTGTTCGGA 2220
 Db 2178 TCAAGTGAAGCTGAAAAAGTGAAGGCCCACTCAAGAAAGCAGAAAGTTTGTTCGGA 2237
 Oy 2221 AGTAACGGATTCATGCTGAAAGCCAAATGCAACAGAAATCTAGCTGTTACGAAATTA 2280
 Db 2238 AGTAACGGATTCATGCTGAAAGCCAAATGCAACAGAAATCTAGCTGTTACGAAATTA 2297
 Oy 2281 TTTGACTCTTCAATTAATGATTAACAATATGATATGTCAGAGCAGAAAAATTAATTCC 2340
 Db 2298 TTTGACTCTTCAATTAATGATTAACAATATGATATGTCAGAGCAGAAAAATTAATTCC 2357
 Oy 2341 GTTGTAAAGGAAGTATCCTTCATCTGTAAGTAAAGAAAAATTAAC 2389
 Db 2358 GTTGTAAAGGAAGTATCCTTCATCTGTAAGTAAAGAAAAATTAAC 2406

RESULT 4
 AAA47604
 ID AAA47604 standard; DNA, 2451 BP.
 XX
 AC AAA47604;
 XX
 DT 20-OCT-2000 (first entry)
 XX
 DE Recombinant variant of Sp36 gene (Sp36A) of *S. pneumoniae*.
 XX
 KW Streptococcus pneumoniae; infection; vaccine; coiled coil region;
 XX hiection triad residue; Sp36; antibody; otitis media;
 KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
 XX meningitis; lobar pneumonia; ds.
 OS Streptococcus pneumoniae.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..2451
 FT /tag= a
 FT /product= Sp36A polypeptide
 XX
 XX W0200037105-A2.

29-JUN-2000.
 21-DEC-1999; 99WO-US30390.
 21-DEC-1998; 98US-0113048.
 (MEDI-) MEDIMUNE INC.
 Johnson LS, Koenig S, Adamou JE;
 WPI: 2000-452129/39.
 P-PSDB; AAB01468.
 Vaccine useful for prophylaxis and treatment of pneumococcal infections
 PT such as otitis media, nasopharyngeal and bronchial infections,
 PT comprises Streptococcus pneumoniae proteins
 PS Disclosure; Page 64-65; 70pp; English.
 XX
 XX Although a number of proteins have been suggested as being involved
 CC in the pathogenicity of Streptococcus pneumoniae, there still remains
 CC a need to identify polypeptides having epitopes in common from
 CC various strains of S. pneumoniae in order to utilise such
 CC polypeptides in vaccines to protect against a wide variety of
 CC S. pneumoniae. New vaccine compositions are described which comprise a
 CC Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino
 CC acids in length that comprise at least one histidine triad residue
 CC (HxxHx) or a coiled-coil region, or an antibody directed against
 CC these features. The vaccine is useful in protecting against infection
 CC by Streptococcus pneumoniae. The vaccine composition comprising
 CC antibodies to be useful for passive immunization for treating
 CC pneumococcal infections which includes otitis media, nasopharyngeal
 CC and bronchial infections.
 CC
 SQ Sequence 2451 BP; 849 A; 467 C; 499 G; 635 T; 1 other;

Query Match 100.0%; Score 2388; DB 21; Length 2451;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAAAGGTTAAGAAATAATCGTGTCTTA 60
 60 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAAAGGTTAAGAAATAATCGTGTCTTA 119
 61 TATAGTGAAGAAACGAGCGCAAAACGAGAAATTTGACTCTCGATGAGGTTAGCA 120
 120 TATAGTGAAGAAACGAGCGCAAAACGAGAAATTTGACTCTCGATGAGGTTAGCA 179
 121 GGTGAAGAAATCAATGCTGAGCAATCGTCATCAAGATACAGACCAAGCTATGTCAC 180
 180 GGTGAAGAAATCAATGCTGAGCAATCGTCATCAAGATACAGACCAAGCTATGTCAC 239
 181 TTCACATGGCGCACTATCATTTATTAATGATGTAAGGTTCTTATGACGCTATCATCG 240
 240 TTCACATGGCGCACTATCATTTATTAATGATGTAAGGTTCTTATGACGCTATCATCG 299
 241 TGAAGAATTCTCATGAAGAATCCAAACTATAAGCTAAAGATGAGGATATGTTAATA 300
 300 TGAAGAATTCTCATGAAGAATCCAAACTATAAGCTAAAGATGAGGATATGTTAATA 359
 301 GGTCAAGGTGATATGTTATCAAGGTAGTGAATAATCTATGTTTAAAGATGTC 360
 360 GGTCAAGGTGATATGTTATCAAGGTAGTGAATAATCTATGTTTAAAGATGTC 419
 361 TGCCCAACGGGATTAAGTCCGTAACAAGAGAAATCAATGACAAAACAAGAGATAG 420
 420 TGCCCAACGGGATTAAGTCCGTAACAAGAGAAATCAATGACAAAACAAGAGATAG 479
 421 TCAACATGTTGAAGGTAAGTCCGTAACAAGAGAAATCAATGACAAAACAAGAGATAG 480
 480 TCAACATGTTGAAGGTAAGTCCGTAACAAGAGAAATCAATGACAAAACAAGAGATAG 539
 481 AGGACGCTATACAGATGATGTTATATCTTAATGCTCTGATATCATAGAGATAC 540

540 AGGACGCTATACAGATGATGTTATATCTTAATGCTCTGATATCATAGAGATAC 599
 541 TGCTGATGCTTATATCGTTCCTTCAAGAGATATACCATTAATCTTCAAGATGAGTT 600
 600 TGCTGATGCTTATATCGTTCCTTCAAGAGATATACCATTAATCTTCAAGATGAGTT 659
 601 ATCAGTACGAGGAGTGGCTGCTGAGAAAGCTCTCTATCTGCTGAGGAAATCTGCAAA 660
 660 ATCAGTACGAGGAGTGGCTGCTGAGAAAGCTCTCTATCTGCTGAGGAAATCTGCAAA 719
 661 TTCAAGAACTATTCGCCGACAAATAGGATTAACATTCACAGAACTGGGTCTTC 720
 720 TTCAAGAACTATTCGCCGACAAATAGGATTAACATTCACAGAACTGGGTCTTC 779
 721 TGTAAAGCAATCCAGAACTACAAATCTTACAGAAAGCAACAGACACTTACAGTCA 780
 780 TGTAAAGCAATCCAGAACTACAAATCTTACAGAAAGCAACAGACACTTACAGTCA 839
 781 AGCAAGTCAAGTAAATGACATGATGATGCTCTTGAAGAGCTCTCAACCTGCTTGA 840
 840 AGCAAGTCAAGTAAATGACATGATGATGCTCTTGAAGAGCTCTCAACCTGCTTGA 899
 841 TCAAGACATGTAGAAATCTGATGCTCTTGTATGATCCAGACAAATCAGAGTCAAC 900
 900 TCAAGACATGTAGAAATCTGATGCTCTTGTATGATCCAGACAAATCAGAGTCAAC 959
 901 AGCTGAGGTGTGAGTCCAGTCCAGACAGAGATATTAACAATTCATCTTCAAT 960
 960 AGCTGAGGTGTGAGTCCAGTCCAGACAGAGATATTAACAATTCATCTTCAAT 1019
 961 GTCGAATTTGAGAGAGCAATGCTGCTGATTAATCCCTGCTTGTGCTCAACATG 1020
 1020 GTCGAATTTGAGAGAGCAATGCTGCTGATTAATCCCTGCTTGTGCTCAACATG 1079
 1021 GGTACCAATTCAGAGCGAGAAACCAACCAAGTCCACCAAGTCCGGAACCTAGTCCAG 1080
 1080 GGTACCAATTCAGAGCGAGAAACCAACCAAGTCCACCAAGTCCGGAACCTAGTCCAG 1139
 1081 CCCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTTGTAGTCACTGCT 1140
 1140 CCCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTTGTAGTCACTGCT 1199
 1141 ACGAAAGTGGGAGAGATATGATGAGAGAAAGGCAATCTCGTATGCTTTC 1200
 1200 ACGAAAGTGGGAGAGATATGATGAGAGAAAGGCAATCTCGTATGCTTTC 1259
 1201 GAAAGATTTACCATCTGAACTGTTAAATCTTGAAGCAAGTATCAAAACAAGAG 1260
 1260 GAAAGATTTACCATCTGAACTGTTAAATCTTGAAGCAAGTATCAAAACAAGAG 1319
 1261 TGTTCACACACTTAACTGCTAAAGAAAGAAATGTTGCTCTGTCGACCAAGATTTTA 1320
 1320 TGTTCACACACTTAACTGCTAAAGAAAGAAATGTTGCTCTGTCGACCAAGATTTTA 1379
 1321 TGATAAGCATATATCTGTTAACTGAGGCTATTAAGCTGTTGAAAATAAGAGTGC 1380
 1380 TGATAAGCATATATCTGTTAACTGAGGCTATTAAGCTGTTGAAAATAAGAGTGC 1439
 1381 TAAATCTGATTTCCAAAGCTTAAAGCAATTTAAGAGCTGAATGATGATCACTTA 1440
 1440 TAAATCTGATTTCCAAAGCTTAAAGCAATTTAAGAGCTGAATGATGATCACTTA 1499
 1441 TAAAGAAATTTGGTATGATTTATTTGGCATTTCTAGACCAATTAATCCATCCAGAG 1500
 1500 TAAAGAAATTTGGTATGATTTATTTGGCATTTCTAGACCAATTAATCCATCCAGAG 1559
 1501 ACTGGCAACCAATTTCTCAATGATGATTAAGAGAGAGTTCGATTCGATTC 1560
 1560 ACTGGCAACCAATTTCTCAATGATGATTAAGAGAGAGTTCGATTCGATTC 1619
 1561 AGCTGATAGTATACAGATGATGTTTATATCTTAATGCTCTGATATCATAGAGATAC 1620

Db 1620 AGCTGATAGTATACACGCTCAGATGTTACATTTTGTATGTAACATGATATATCATGAGCA 1679
 Qy 1621 TGAAGAGATGATATGTAAGCCCTCATATGGGCGCATAGTCACTGGATTTGGAAAAAGATAG 1680
 Db 1680 TGAAGAGATGATATGTAAGCCCTCATATGGGCGCATAGTCACTGGATTTGGAAAAAGATAG 1739
 Qy 1681 CTTTCTGATATGAAAAAGTTGCAAGCTCAAGCTTATCTAAAGAAAAAGGATCTTACC 1740
 Db 1740 CTTTCTGATATGAAAAAGTTGCAAGCTCAAGCTTATCTAAAGAAAAAGGATCTTACC 1799
 Qy 1741 TCCATCTCAGACGAGATGTTAAAGCAATCCAACTGAGATATGACAGCAGCTATTTA 1800
 Db 1800 TCCATCTCAGACGAGATGTTAAAGCAATCCAACTGAGATATGACAGCAGCTATTTA 1859
 Qy 1801 CAATGCTGGAAGGAAAAAGAAATCCATCGTTGAGCTTCAATATATGTTGAGCA 1860
 Db 1860 CAATGCTGGAAGGAAAAAGAAATCCATCGTTGAGCTTCAATATATGTTGAGCA 1919
 Qy 1861 TACAGTTGAGGTTAAAAAGGTAATTTGATTAATTCCTCATTAAGATCATTAACATATAT 1920
 Db 1920 TACAGTTGAGGTTAAAAAGGTAATTTGATTAATTCCTCATTAAGATCATTAACATATAT 1979
 Qy 1921 TAAATTTGCTGTTGTTGATGATCACAATCAAGCTCCAAATGGCTTATACCTTGAAGA 1980
 Db 1980 TAAATTTGCTGTTGTTGATGATCACAATCAAGCTCCAAATGGCTTATACCTTGAAGA 2039
 Qy 1981 TTTGTTGCGAGATTAAGTACTAGTGAACACCTGACGAAGCTCCAACTTCAATGA 2040
 Db 2040 TTTGTTGCGAGATTAAGTACTAGTGAACACCTGACGAAGCTCCAACTTCAATGA 2099
 Qy 2041 TGGATGGGCAATGCGCAGTGAAGCATGTGTAGGCAAGAAAGCAACAGTGAAGATCCAA 2100
 Db 2100 TGGATGGGCAATGCGCAGTGAAGCATGTGTAGGCAAGAAAGCAACAGTGAAGATCCAA 2159
 Qy 2101 TAAAGACTTCAAAAGCGGATGAAGAGCCAGTGAAGAAACACTGCTGAGCCAGAAATGCC 2160
 Db 2160 TAAAGACTTCAAAAGCGGATGAAGAGCCAGTGAAGAAACACTGCTGAGCCAGAAATGCC 2219
 Qy 2161 TCAAGTAGAGCTGAAAAAGTAGAGCCCACTCAAGAGAGAAAGTTTGTCTGGGAA 2220
 Db 2220 TCAAGTAGAGCTGAAAAAGTAGAGCCCACTCAAGAGAGAAAGTTTGTCTGGGAA 2279
 Qy 2221 AGTAAAGGATCTAGCTGAAAAAGCCCAATGCAACAGAAACTAGCTGTTTACGAAATTA 2280
 Db 2280 AGTAAAGGATCTAGCTGAAAAAGCCCAATGCAACAGAAACTAGCTGTTTACGAAATTA 2339
 Qy 2281 TTTGACTCTTCAAAATTTATGAGTAACAATATGATCTGCGAGAGAGCAAAAAATTTCTGC 2340
 Db 2340 TTTGACTCTTCAAAATTTATGAGTAACAATATGATCTGCGAGAGAGCAAAAAATTTCTGC 2399
 Qy 2341 GTTGTAAAGAGATATCTCTCATCTGTAAGTAAAGAAAAAATTAAC 2389
 Db 2400 GTTGTAAAGAGATATCTCTCATCTGTAAGTAAAGAAAAAATTAAC 2448

RESULT 5
 AAV52227
 ID AAV52227 standard; DNA; 8195 BP.

AC AAV52227;
 XX 23-OCT-1998 (first entry)

DE Streptococcus pneumoniae genome fragment SEQ ID NO:94.
 XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
 KW computer readable medium; vaccine; pharmaceutical composition; ds.
 XX Streptococcus pneumoniae.
 XX MO9818931-A2.
 XX 07-MAY-1998.

PD 07-MAY-1998.

XX 30-OCT-1997; 97WO-US19588.
 PF 31-OCT-1996; 96US-0029960.
 BR (HUMA-) HUMAN GENOME SCI INC.
 PA Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
 PI Kunsch CA, Rosen CA;
 XX WPI; 1998-272225/24.
 DR Computer-readable medium with recorded Streptococcus pneumoniae
 PT polynucleotide sequences - useful in diagnostic kits and assays, and
 PT pharmaceutical compositions and vaccines for Streptococcus
 PT pneumoniae
 PS Claim 1; Page 727-732; 1409pp; English.
 CC The present invention describes a computer readable medium which has
 CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
 CC recorded on it, or a representative fragment or a sequence at least 95%
 CC identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in
 CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
 CC Streptococcus pneumoniae. The present invention also describes an
 CC isolated nucleic acid molecule encoding a homologue of any of the
 CC fragments of the S. pneumoniae genome (SEQ ID NO:1 to 391) where the
 CC nucleic acid molecule is produced by a process comprising: (a) screening
 CC a genomic DNA library using as a probe a target sequence defined by any
 CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
 CC library which contain sequences that hybridize to the target sequence and
 CC isolating the nucleic acid molecules from the members; or (b) isolating
 CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
 CC molecules whose nucleotide sequence is homologous to amplification
 CC primers derived from the fragment of the S. pneumoniae genome to prime
 CC the amplification and isolating the amplified sequences. The computer
 CC readable medium can be used in a computer-based system for identifying
 CC fragments of the S. pneumoniae genome of commercial importance, or
 CC expression modulating fragments of the S. pneumoniae genome. Products
 CC from the present invention can be used in diagnosis kits and assays, and
 CC pharmaceutical compositions and vaccines for S. pneumoniae.
 CC
 SQ Sequence 8195 BP; 2688 A; 1622 C; 1777 G; 2105 T; 3 other;
 Query Match 100.0%; Score 2388; DB 19; Length 8195;
 Best Local Similarity 100.0%; Pred. No. 0; Mismatches 0; Gaps 0;
 Matches 2388; Conservative 0; Indels 1; Indels 0; Gaps 0;
 Qy 1 TTTCTTACAGTTGGAGCTGATATCAAGCTAGAACGTTAAGAAAAATATCGTGTCTTA 60
 Db 3053 TTTCTTACAGTTGGAGCTGATATCAAGCTAGAACGTTAAGAAAAATATCGTGTCTTA 3112
 Qy 61 TATATGATGAAAAACAGCGAGCGCAAAAAAGGAGAAATTTGACTCTGATAGGTTAGCA 120
 Db 3113 TATATGATGAAAAACAGCGAGCGCAAAAAAGGAGAAATTTGACTCTGATAGGTTAGCA 3172
 Qy 121 GCGTGAAGAGATCAATCTGAGCAAAATCGTATCAAGTAAACAGCAAGGCTATGTCAC 180
 Db 3173 GCGTGAAGAGATCAATCTGAGCAAAATCGTATCAAGTAAACAGCAAGGCTATGTCAC 3232
 Qy 181 TTCACATGCGACCACTATCATTTATTAACAATGTAAGTTCTTATGACGCTATCATCAG 240
 Db 3233 TTCACATGCGACCACTATCATTTATTAACAATGTAAGTTCTTATGACGCTATCATCAG 3292
 Qy 241 TGAAGAAATTAATCTATGAAAGATCCAAATCAATTAAGCTTAAAGATGAGATATTTGAATGA 300
 Db 3293 TGAAGAAATTAATCTATGAAAGATCCAAATCAATTAAGCTTAAAGATGAGATATTTGAATGA 3352
 Qy 301 GGTCAAGGCTGAGATATTTATCAAGTATGAGAAATTAATCTTATACCTTAAGATGC 360
 Db 3353 GGTCAAGGCTGAGATATTTATCAAGTATGAGAAATTAATCTTATACCTTAAGATGC 3412
 Qy 361 TGCCACGCGGATTAACGTCGTACAAAGAGAAATCAATCGACAAACAAAGACATAG 420

Db 3413 TCCCAACGGGATTAAGCTCCGTACAAAAGAGAAATCAACAAAACAAGAGCATAG 3472
 Qy 421 TCACATCGTGAAGTGAACCTCCAGAAAACGATGCTGCTTGGCCACGCTTCGA 480
 Db 3473 TCACATCGTGAAGTGAACCTCCAGAAAACGATGCTGCTTGGCCACGCTTCGA 3532
 Qy 481 AGGACGCTATACATAGATGATGCTTATCTTTAAATGCTTCGATATCATAGAGATAC 540
 Db 3533 AGGACGCTATACATAGATGATGCTTATCTTTAAATGCTTCGATATCATAGAGATAC 3592
 Qy 541 TGGTATGCTTATATGCTTCCTCATGAGATCATTAACATTAATTCCTAAGATAGTT 600
 Db 3593 TGGTATGCTTATATGCTTCCTCATGAGATCATTAACATTAATTCCTAAGATAGTT 3652
 Qy 601 ATCAGTAGAGATGCTGCTGCGACAAAGCTTCCTATCTGCTGCGAGAAATCTGCACA 660
 Db 3653 ATCAGTAGAGATGCTGCTGCGACAAAGCTTCCTATCTGCTGCGAGAAATCTGCACA 3712
 Qy 661 TTCAGAACTTATCCGCGACAAAATAGCATTAACCTTCAAGAACAACTGGGTACTTC 720
 Db 3713 TTCAGAACTTATCCGCGACAAAATAGCATTAACCTTCAAGAACAACTGGGTACTTC 3772
 Qy 721 TGTAAAGCAATCCAGGAATCAAAATCTAACACAGCAACACAGCAACATPACAGTCA 780
 Db 3773 TGTAAAGCAATCCAGGAATCAAAATCTAACACAGCAACACAGCAACATPACAGTCA 3832
 Qy 781 ACCAAGTCAAAATGAATGATGATGCTCTTGAACAGCTCTCAAACTGCTTGAAG 840
 Db 3833 ACCAAGTCAAAATGAATGATGATGCTCTTGAACAGCTCTCAAACTGCTTGAAG 3892
 Qy 841 TCAAGGACATGATAGATGATGCTGCTTGAACAGCTCTCAAACTGCTTGAAG 900
 Db 3893 TCAAGGACATGATAGATGATGCTGCTTGAACAGCTCTCAAACTGCTTGAAG 3952
 Qy 901 AGCTAAGGCTGTGACGTCGACACAGAGATCATTAACCACTTATCTCTCAAT 960
 Db 3953 AGCTAAGGCTGTGACGTCGACACAGAGATCATTAACCACTTATCTCTCAAT 4012
 Qy 961 GCTCTAATTTGGAAGAACGATGCTGCTGCTTATCTCTGCTTATCTGCTTCAACATTC 1020
 Db 4013 GCTCTAATTTGGAAGAACGATGCTGCTGCTTATCTCTGCTTATCTGCTTCAACATTC 4072
 Qy 1021 GGTACAGATTCAGAGCCAGAACCAACAGTCCACAAACGAGCTCCGGAACCTAGTCAGG 1080
 Db 4073 GGTACAGATTCAGAGCCAGAACCAACAGTCCACAAACGAGCTCCGGAACCTAGTCAGG 4132
 Qy 1081 CCCGAACTGACCAAAATCTTAAATAGACTCAAAATCTCTGCTTATGCTAGCTGCT 1140
 Db 4133 CCCGAACTGACCAAAATCTTAAATAGACTCAAAATCTCTGCTTATGCTAGCTGCT 4192
 Qy 1141 AGGAAAAGTTGGGGAAGGATATGATTCGAGAAAAGGCAATCTCTGCTTATGCTTTC 1200
 Db 4193 AGGAAAAGTTGGGGAAGGATATGATTCGAGAAAAGGCAATCTCTGCTTATGCTTTC 4252
 Qy 1201 GAAAGATTACCATCTGAACCTGTTAAATCTTGAAGCAAGTTATCAAAACAAGAGAG 1260
 Db 4253 GAAAGATTACCATCTGAACCTGTTAAATCTTGAAGCAAGTTATCAAAACAAGAGAG 4312
 Qy 1261 TGTTCACACACTTAACTGCTTAAATAGAAAATGTTGCTCTGTCGACCAAGAAATTTA 1320
 Db 4313 TGTTCACACACTTAACTGCTTAAATAGAAAATGTTGCTCTGTCGACCAAGAAATTTA 4372
 Qy 1321 TGTAAAGATATATCTGTTAACTGAGGCTCATTAAGCTGTTGAAATTAAGGCTG 1380
 Db 4373 TGTAAAGATATATCTGTTAACTGAGGCTCATTAAGCTGTTGAAATTAAGGCTG 4432
 Qy 1381 TAAATCTGATTCAGAGCTTAGACAAATTAATTAAGAGCTGAATGATGATCACTAA 1440
 Db 4433 TAAATCTGATTCAGAGCTTAGACAAATTAATTAAGAGCTGAATGATGATCACTAA 4492
 Qy 1441 TAAAGAAAATTTGATAGATGATTTATGCTTCTAGACCAATTAACCATCCAGAGCG 1500

Db 4493 TAAAGAAAATTTGATAGATGATTTATGCTTCTAGACCAATTAACCATCCAGAGCG 4552
 Qy 1501 ACTTGGCAAAACCAATTTCTCAATGATGATCTGAAGCGAAGTTCGATTCCTCAAT 1560
 Db 4553 ACTTGGCAAAACCAATTTCTCAATGATGATCTGAAGCGAAGTTCGATTCCTCAAT 4612
 Qy 1561 AGCTGATAAGTTAGAACGCTGATGATGCTTATCTTTGATGAACATGATATATCACTGA 1620
 Db 4613 AGCTGATAAGTTAGAACGCTGATGATGCTTATCTTTGATGAACATGATATATCACTGA 4672
 Qy 1621 TGAAGAGATGATATGTAACGCTCATATGCGCCATATGCTCATGATTTGAAAAATAG 1680
 Db 4673 TGAAGAGATGATATGTAACGCTCATATGCGCCATATGCTCATGATTTGAAAAATAG 4732
 Qy 1681 CCTTCTGATAGAAAAATGCGCTCAGCTCAAGCTTATCTAAAGAAAAAGTATCTTACC 1740
 Db 4733 CCTTCTGATAGAAAAATGCGCTCAGCTCAAGCTTATCTAAAGAAAAAGTATCTTACC 4792
 Qy 1741 TCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGGAGATAGTCAGAGCTATTTA 1800
 Db 4793 TCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGGAGATAGTCAGAGCTATTTA 4852
 Qy 1801 CAATGCTGTGAAGGGGAAAAAGAAATTCAGTCTGCTTCAATATATGCTTGAGCA 1860
 Db 4853 CAATGCTGTGAAGGGGAAAAAGAAATTCAGTCTGCTTCAATATATGCTTGAGCA 4912
 Qy 1861 TACAGTTAGGTTAAACCGGTAATTTGATTTCTCTATAGATCATTAACATTAAT 1920
 Db 4913 TACAGTTAGGTTAAACCGGTAATTTGATTTCTCTATAGATCATTAACATTAAT 4972
 Qy 1921 TAAATTTGCTGTTGATGATCACAATCAAACTCCAAATGCTATACCTTGGAAGA 1980
 Db 4973 TAAATTTGCTGTTGATGATCACAATCAAACTCCAAATGCTATACCTTGGAAGA 5032
 Qy 1981 TTTGTTTGGAGATTAAGTACTAGTGAACACCCCTGACGAACTTCATTAATGA 2040
 Db 5033 TTTGTTTGGAGATTAAGTACTAGTGAACACCCCTGACGAACTTCATTAATGA 5092
 Qy 2041 TGGATGGGCAATGCGAGTGAACATGCTTGAAGCAAGAACCAAGTGAAGATTCAAA 2100
 Db 5093 TGGATGGGCAATGCGAGTGAACATGCTTGAAGCAAGAACCAAGTGAAGATTCAAA 5152
 Qy 2101 TAAAGATTCAAAGGGGATGAAGAGCCAGTAGAGAAAACCTGCTGAGCCGAAGTCCC 2160
 Db 5153 TAAAGATTCAAAGGGGATGAAGAGCCAGTAGAGAAAACCTGCTGAGCCGAAGTCCC 5212
 Qy 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGAGCAAGATTTTGCGGA 2220
 Db 5213 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGAGCAAGATTTTGCGGA 5272
 Qy 2221 AGTAAAGGATTTAGTCTGAAGAGCCAAATGCAACAGAACTTACGCTGTTAGCAATTA 2280
 Db 5273 AGTAAAGGATTTAGTCTGAAGAGCCCAATGCAACAGAACTTACGCTGTTAGCAATTA 5332
 Qy 2281 TTTGACTCTTCAAAATTAATGATTAACATAGATCATGACAGAGAGCAAAAAATTTCTGCC 2340
 Db 5333 TTTGACTCTTCAAAATTAATGATTAACATAGATCATGACAGAGAGCAAAAAATTTCTGCC 5392
 Qy 2341 GTTGTAAAGAGATTAATCTTCACTGTAAGTAAAGAAAAATTAAC 2389
 Db 5393 GTTGTAAAGAGATTAATCTTCACTGTAAGTAAAGAAAAATTAAC 5441

RESULT 6
 ABS56454/C
 ID ABS56454 standard; DNA; 2162598 BP.
 AC ABS56454;
 XX
 XX 10-FEB-2003 (first entry)
 XX
 DE Streptococcus pneumoniae type 4 strain complete genome.


```

Db 1005750 GAAAATTTACATCTGAAAACGTTAAAAATCTGAAACCAAGTTATCAAAAAGAGAG 1005691
QY 1261 TGTTCACACACTTAACTGCTAAAAAAGAAATGTGCTCTCGGACCAAGAAATTTTA 1320
Db 1005690 TGTTCACACACTTAACTGCTAAAAAAGAAATGTGCTCTCGGACCAAGAAATTTTA 1005631
QY 1321 TGATTAAGCATTAATCTGTAACTGAGGCTCATAAAGCTTGTGNAATTAAGGCTCG 1380
Db 1005630 TGATTAAGCATTAATCTGTAACTGAGGCTCATAAAGCTTGTGNAATTAAGGCTCG 1005571
QY 1381 TAATTCGATTTCCAGCTTGAACAATTTAGAACGCTTGAATGATGATGCACTPA 1440
Db 100570 TAATTCGATTTCCAGCTTGAACAATTTAGAACGCTTGAATGATGATGCACTPA 1005511
QY 1441 TAAAGAAAAATGTGATGATTAATTTGSCATTTCTGACAGCAATTAACCATCCAGAGG 1500
Db 1005510 TAAAGAAAAATGTGATGATTAATTTGSCATTTCTGACAGCAATTAACCATCCAGAGG 1005451
QY 1501 ACTTGCAAAACCAATTTCTCAATTTAGATATCTGAAGACGAAGTTGCTATTGCTCAAT 1560
Db 1005450 ACTTGCAAAACCAATTTCTCAATTTAGATATCTGAAGACGAAGTTGCTATTGCTCAAT 1005391
QY 1561 AGCTGATAAGTATCAACGCTCAGATGTTTCAATTTTGTGAACATGATATATCATGTA 1620
Db 1005390 AGCTGATAAGTATCAACGCTCAGATGTTTCAATTTTGTGAACATGATATATCATGTA 1005331
QY 1621 TGAAGGAGATGATATGTAACCCCTCATATGGGCCATATCTGATTTGAAAAAGATAG 1680
Db 1005330 TGAAGGAGATGATATGTAACCCCTCATATGGGCCATATCTGATTTGAAAAAGATAG 1005271
QY 1681 CCTTCTGATAAGAAAAAGTTGCAGCTCAAGCTATATCTAAAGAAAAAGATCTCTACC 1740
Db 1005270 CCTTCTGATAAGAAAAAGTTGCAGCTCAAGCTATATCTAAAGAAAAAGATCTCTACC 1005211
QY 1741 TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGAGATAGTGCAGACGATTTTA 1800
Db 1005210 TCCATCTCCAGACGAGATGTTAAAGCAATCCAACTGAGATAGTGCAGACGATTTTA 1005151
QY 1801 CAATGCTGTAAGGAGGAAAAAGCAATTCCTGCTGCTTCCATATATGTTGAGGA 1860
Db 1005150 CAATGCTGTAAGGAGGAAAAAGCAATTCCTGCTGCTTCCATATATGTTGAGGA 1005091
QY 1861 TACAGTTGAGTTAAAAAGCGTAATTTGATATTCCTCATAGAGATCATCATATAT 1920
Db 1005090 TACAGTTGAGTTAAAAAGCGTAATTTGATATTCCTCATAGAGATCATCATATAT 1005031
QY 1921 TAAATTTGCTGTTGATGATCACAACATACAAAGCTCCAAATGGCTATACCTTGGAGA 1980
Db 1005030 TAAATTTGCTGTTGATGATCACAACATACAAAGCTCCAAATGGCTATACCTTGGAGA 1004971
QY 1981 TTTTGTTCGAGATTAAGTACTAGTAGAACAACCTGACGAAGCTCCACATTTCAATGA 2040
Db 1004970 TTTTGTTCGAGATTAAGTACTAGTAGAACAACCTGACGAAGCTCCACATTTCAATGA 1004911
QY 2041 TGGATGGGGCAATGCGACGATGTTAGTGAAGAAAGCCACAGGAAGATCCAAA 2100
Db 1004910 TGGATGGGGCAATGCGACGATGTTAGTGAAGAAAGCCACAGGAAGATCCAAA 1004851
QY 2101 TAAAGACTTCAAAAGCGATGTAAGAGCCAGTAGAGAAAACACTGCTGAGCCAGAACTCC 2160
Db 1004850 TAAAGACTTCAAAAGCGATGTAAGAGCCAGTAGAGAAAACACTGCTGAGCCAGAACTCC 1004791
QY 2161 TCAAGTGAAGATGAAAAAGTGAAGCCCACTCAAGAGACGAAGATTTTCTTGGCGAA 2220
Db 1004790 TCAAGTGAAGATGAAAAAGTGAAGCCCACTCAAGAGACGAAGATTTTCTTGGCGAA 1004731
QY 2221 AGTAAAGGATTTAGTCTGAAGCCCAATGCAACAGAACTTAGAGGTTTACGAATPA 2280
Db 1004730 AGTAAAGGATTTAGTCTGAAGCCCAATGCAACAGAACTTAGAGGTTTACGAATPA 1004671
QY 2281 TTTGACTCTTCAAAATTAAGATTAACATATGATCAATGGCAGAAAGCAAAAAATTAATTGC 2340

```

```

Db 1004670 TTTGACTCTTCAAAATTAAGATTAACATATGATCAATGCGACAGAAAGAAAAATTAATTGC 1004611
QY 2341 GTTGTAAAGAGATATCCCTCATCTGTAGTAGAGAAAAAATPAAC 2389
Db 1004610 GTTGTAAAGAGATATCCCTCATCTGTAGTAGAGAAAAAATPAAC 1004562

RESULT 7
AAA65731
ID AAA65731 standard; DNA; 2523 BP.
XX
XX AAA65731;
AC
AC 21-NOV-2000 (first entry)
DT
DT Streptococcus pneumoniae BVH-11 gene SEQ ID NO:3.
DE
DE Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal; ds.
OS Streptococcus pneumoniae.
XX
XX WO20039299-A2.
XX
XX 06-JUL-2000.
XX
XX 20-DEC-1999; 99WO-CA01218.
XX
XX 23-DEC-1998; 98US-0113800.
XX
XX (BIOC-) BIOCHEM PHARMA INC.
XX
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
XX DR P-PSDB; AAB12716.
XX
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteraemia and/or pneumonia -
XX
XX Example 2; Fig 3; 106pp; English.
XX
XX The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens
XX have bactericidal activity. The nucleic acids, encoding the protein
XX antigens, may be used for the recombinant production of the proteins
XX they encode. The protein antigens may then be used as vaccines for the
XX prevention and treatment of Streptococcal infections in mammals
XX (especially humans) which result in, e.g. meningitis, otitis media,
XX bacteraemia and/or pneumonia. The present sequence encodes the
XX S. pneumoniae BVH-11 protein antigen.
XX
XX Sequence 2523 BP; 879 A; 523 C; 526 G; 595 T; 0 other;

Query Match 57.5%; Score 1374.2; DB 21; Length 2523;
Best Local Similarity 73.9%; Pred. No. 0;
Matches 1919; Conservative 0; Mismatches 539; Indels 102; Gaps 2;

1 TTTCTACAGTTGGAGCTGATTCAGAGCGTTAGGAAAAATATCGGTTTCTTA 60
Db 60 TGTCTATATACTAGTTTGCATCAAGCTCAAACTGTAAAAAATAATCGTTTCTTA 119
QY 61 TATAGATGAAAAACAAGCGACGCAAAAAAGAGAAATTTGACTCTGATGAGTTAGCA 120
Db 120 TATAGATGAAAAACAAGCGACGCAAAAAAGAGAAATTTGACTCTGATGAGTTAGCA 179
QY 121 GCGTGAAGAGATTCATGCTGAGCAAAATGTGATCAAGATACAGACCAAGCTATGTCAC 180
Db 180 GCGTGAAGAGATTCATGCTGAGCAAAATGTGATCAAGATACAGGTTATGTCAC 239
QY 181 TTTCAATGCGACGACATATCATTTATCAATGTAAGGTTCTTATGACGATCATGAC 240

```

Db 240 CTCCTCATGGAGACCATTTATCATTTACTAATATGCGAAGGCCCTTATGATGCCATCATAG 299
 Qy 241 TGAAGATTTACTCATGAAGAATCCAAACATATAGCTAAAGATGAGATTTGTTAATGA 300
 Db 300 TGAAGAGCTCTCCATGAAAAGATCCGAATTTATCATGTTGAAGATTCAGACATTTGTAAGA 359
 Qy 301 GGTCAAGGGTGGATATGTTATCAAGGTAGTGAATAATATATGTTTACCTTTAAGATGC 360
 Db 360 AATCAAGGGTGGTATGTCATTTAAGGTAAACGTAAATATCTATGTTTACCTTTAAGATGC 419
 Qy 361 TGCCCAACGGGATTAACGTCGTCACAAAAGAGAAATCAATGCAACAAAACAAAGCATAG 420
 Db 420 AGCTCATGCGGATTAATGTCGTCACAAAAGAGAAATCAATGCGCAAAAACAAAGAACTAG 479
 Qy 421 TCAACATGCTGAAGGTGGAACCTCAAGAAAGATGTCGTGTTGCTTGCGCATGCTTGCA 480
 Db 480 TCAGATGCTGAAGAGAGGAGCTTCAGCAAAACGATGCTGGGAGAGCTTTGACGTTCA 539
 Qy 481 AGGACGCTATACATGATGATGTTAATCTTTAATGCTTGTATATCATAGAGATAC 540
 Db 540 GGGAGCTTACACACAGATGATGTTATATCTTCAATGATCTGATATCATGAGATAC 599
 Qy 541 TGTGATGCTTATATATGTTCTCTCATGAGATCATTAACATTAATTCCTTAAGATGAT 600
 Db 600 GGGCATGCTTATATGTTCTCTCATGAGATCATTAACATTAATTCCTTAAGATGAT 659
 Qy 601 ATCACTAGCGAGTTGGCTGTCAGAAAGCCCTTCTATCTGTCAGAGAAATCTGTCAA 660
 Db 660 ATCACTAGCGAGTTGGCTGTCAGAAAGCCCTTCTATGCTGCGGAAATCTGTCAA 719
 Qy 661 TTCAAGAACTTATCCCGCAAAAATAGCATTAACCTTCAAGAAACCTGGGTACCTTC 720
 Db 720 TTTAAGAACTTATCCCGCAAAAATAGCATTAACCTTCAAGAAACCTGGGTACCTTC 779
 Qy 721 TGTAAAGATCCAGAACTACAAATCTAACACAGCAACACAGCAACCTAACAGTCA 780
 Db 780 TGTAAAGATCCAGAACTACAAATCTAACACAGCAACACAGCAACCTAACAGTCA 839
 Qy 781 AGCAAGTCAAAATATGATCATGATAGTCTCTTGAACAGCTCTCAAACTGCCCTTAG 840
 Db 840 AGCAAGTCAAAATATGATCATGATAGTCTCTTGAACAGCTCTCAAACTGCCCTTAG 899
 Qy 841 TCAACGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 Db 900 TCAACGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 959
 Qy 901 AGCTAGAGGTGTTGAGTGCACACAGGATCATTAACATTCCTTACTTCAAT 960
 Db 960 CGCCAGAGGTGTTGAGTGCCTCATGATGATTAACATTCCTTACTTCAAT 1019
 Qy 961 GTCTGAATTTGAGAAAGCAATGCTCTGTTATTTCCCTTGTATGCTTCAACCATTC 1020
 Db 1020 GTCTGAATTTGAGAAAGCAATGCTCTGTTATTTCCCTTGTATGCTTCAACCATTC 1079
 Qy 1021 GGTACGATTTCAAGGCGAGAAACAAAGTCCACACCGACTCCGGAACCTAGTCCAG 1080
 Db 1080 GGTACGATTTCAAGGCGAGAAACAAAGTCCACACCGACTCCGGAACCTAGTCCAG 1139
 Qy 1081 CCGGCAACCTGACCAAAATCTTAAATAGACTCA-----ATTCTTCTTGGT 1128
 Db 1140 TCCGCAACCTGACCAAAATCTTCAACAGCTCCAGCAATCCAAATTTGATGAGAAATTTGGT 1199
 Qy 1129 TAGTACGCTGTTACGAAAAGTTGGGAGAGATATGTTTCAAGAAAAGGGCATCTTCG 1188
 Db 1200 CAAGAACCTGTTCCAAAAGTAGGCGATGTTATGCTTGTGAGGAAATGAGGTTTCG 1259
 Qy 1189 TTATGCTTTTGGGAAGATTTTACATCTGAAACTGTTAAATCTTGAAGCAAGTTATC 1248
 Db 1260 TTATATCCAGCAAGATCTTTCAAGCAAAACAGCAGAGCATTTGATAGAACTGGC 1319
 Qy 1249 AAAACAGAGATGTTTCAACACTTTAACTGCTAATAAAAGAAAATGTTGCTCTGCTGA 1308
 Db 1320 CAAGCAGGAAGTTATCTCATATAGCTAGAGACTAAGAAAATGACCTCCCTTAGTGA 1379

Qy 1309 CCAAGATTTTATGATTAAGCATATATCTGTTACTGAGGCTCATTAAGCCTTGTGN 1368
 Db 1380 TCGAATTTTACATTAAGCTTATGACTTACTAGCAAGAAATTCACCAAGTTTACTTGA 1439
 Qy 1369 AAATTAAGGCTGTAATTCGATTTTCAAGCCTTAGACAAATTAATTAAGCGCTGAATGA 1428
 Db 1440 TAATTAAGGCTGTAAGCTGTTTGAAGCTTTGATGATTAACCTGTTGAACGACTCAAGGA 1499
 Qy 1429 TGAATGACTTAATAAGAAAATTTGATGATGATTTATTTGGCATCTCTAGACCAATTAC 1488
 Db 1500 TGCTCAAGTATTAAGTCAAGTTAGTGAATATTTCTGCTTGTAGCTTCGATTCG 1559
 Qy 1489 CCATCCAGAGCACTTGCAAAACCAATTCCTCAATGAGTATPACTGAAGCAAGTTTCG 1548
 Db 1560 TCATCCAGAAAGTTTGAAGAAACCAATTCGCAAAATTAATCTACACTATGATGATTTCA 1619
 Qy 1549 TATGCTCAATTAAGCTGATTAAGTATACACGTCAGATGTTTACATTTTGAAGAACATGA 1608
 Db 1620 AGTAGCAAGTTGGCAGCAAGTACACACAGAAAGACGGTTATATCTTGTATCTCGTGA 1679
 Qy 1609 TATATCAAGTATGAAGAGATGATATGTAACGCTCATATGAGGCCATAGCTAGT 1668
 Db 1680 TATTAACAGTATGAGGGGAGTGCCTATGTAACCTCAATATGACCAATAGCCACTGAT 1739
 Qy 1669 TGAAGAAATATGACCTTTCTGATTAAGAAAAGTTGACGCTCAAGCTATTAAGAAA 1728
 Db 1740 TAAAGAAATATGTTTGTCTGAGCTGAGAGACGCGACCCAGCTTATGCTTAAGAGAA 1799
 Qy 1729 AGGTATCCATCTTCATCTCCAGACGAGATGTTTAAAGCAATCAACTGAGATAGTGC 1788
 Db 1800 AGTTTGAACCCCTCTTCGACAGACATCAGATTCAGAAATACTGAGGCAAAAGAGC 1859
 Qy 1789 AGCAGTATTTAAGATGCTGTGAAGAGGGAAGAAACGAATTCATCTGTCGACTTCATA 1848
 Db 1860 AGAAGCTATCTACAAACCGGGAAGAGCTTAAGAGGTGCACCTTATGTAAGCTTGA 1919
 Qy 1849 TATGTTGAGCATACAGTGAAGTTAAAGAGGTAATTTGATTTTCTCTTAAGATCA 1908
 Db 1920 CATCTTCAATATATCTGTAAGTCAAAGAGGTAATTTGATTTTCTCTTAAGATCA 1979
 Qy 1909 TTACCATATATTAATTTGCTTGTGTTGATGATCACATACAAAGCTCCAAATGCTGA 1968
 Db 1980 TTACCATATATTAATTTGCTTGTGTTGATGATGATGATGATGATGATGATGATGAT 2039
 Qy 1969 TACCTTGAAGATTTGTTGCAAGATTTAAGTATACGTAGAACACCTTGACGAAGCTCC 2028
 Db 2040 TACTCTGAGATCTTTTGGGACGTGCAAGTACTATGTCGAAATCCAAACGAAGCTCC 2099
 Qy 2029 ACATTTAATGATGATGAGGCAATGCAAGGATGATGTTAGCAAGAAAGCAACAG 2088
 Db 2100 GCATTCAGATATATGTTTGTGTAAGCTGAGCAACATGTTCAAGAAACAAATTTGCTCA 2159
 Qy 2089 TGAAGATCAAAATTAAGAACTTCAAGCGGATGAGAG----- 2125
 Db 2220 GGAAGAAACCCCTGAGAGAGAAACCAAGCGAAGAAACCAAGTCTCCAAAACCAAC 2279
 Qy 2216 -----CAAGTAGAGAAACACTGCTGAGCCAGAAAGTCCCTCAAGTAGAGACTGAAA 2178
 Db 2280 AGAGAAACCAAGAAAGATATCAAGAGAAATCAAGAACTTGAAGTCCGAGCTGAAA 2339
 Qy 2179 AGTAGAAGCCCAACTCAAGAGAGAGAAATTTGCTTGCAGAAATTAAGAAATTTAGTCT 2238
 Db 2340 GGTGAGAGAAATCAAGAGAGGCTGAAGATTTACTTGAAATTAACAGATTCGAATAT 2399
 Qy 2239 GAAAGCAATGCAACAGAACTTACGCTGTTTACGAATTAATTTGACTCTTCAATTTAT 2298
 Db 2400 CAAGTCAATGCTCAAGAGACTTCAAGAGATTAATAATTAATTTACTATTTGACACCA 2459


```

Db      1424 TCGAAGATTTTACAAATAGGCTTATGACTTACTACGAAAGAAATTCACCAAGATTTCCTTGA 1483
Oy      1369 AATAAGGCTGTAATTCGATTTTCCAAAGCCTTAGACAAATATTAGAACGCTTGATGA 1428
Db      1484 TAAATAAGGTGACAAAGTTGATTTTGAGGCTTTSATTAACCTGTGGAAACGACTAAGA 1543
Oy      1429 TGAATCGACTAATAAGAAAAATGGTAGATGATTTATTTGGATTCCTAGCAACCAATTAC 1488
Db      1544 TGTCGAAGTGAATAAGTCAAGTAGTGATGATATTCCTGGCTTCTTAGCTCCGATTGCG 1603
Oy      1489 CCATCCAGAGCACTGGCCAAACCAATTTCTCAATTGAGTACTGAAAGAGAAATTGCG 1548
Db      1604 TCATCCAAAGCGTTTAGAAAAACCAATGCCCAATTCCTACACTGATGAGAGATTCGA 1663
Oy      1549 TATTGCTCAATTAGCTGTAAGTATACACGTCAGATGTTTACATTTTGTGATGAACAG 1608
Db      1664 AGTAGCCAAAGTTGGACGACGACATACACACAGAGAACGCTTATATCTTTGATCCTCGGA 1723
Oy      1609 TATAATCAGTGTAGAGAGATGATGATGTAAAGCCTCATATGGGCCATAGTCACTGAT 1668
Db      1724 TATAACGAGTGTAGAGGGGATGCTATGTAACTCAATATGACCCATAGCCACTGGAT 1783
Oy      1669 TGGAAAAATGACCTTTCTGATAGAAAAAGTTGCAGCTCAAGCTTACTTAAAGAAA 1728
Db      1784 TAAAAAAGATAGTTGTCTGAAGCTGAGAGACGGCAGCCGCTTATGCTAAAGAGAA 1843
Oy      1729 AGGTATCCTACCTCCATCCAGACGACAGATGTTAAAGCAATCCAACTGGAGATAGTC 1788
Db      1844 AGGTTTGACCCCTCTTGACAGACCATCAGGATTCAGGAAATCTGAGGCAAAAGAGC 1903
Oy      1789 AGCAGCTATTTACAAATCGTGTGAAAAGGGGAAAAAGCAATTCGACTGTTGCACTTCATA 1848
Db      1904 AGAAGCTATCTACAAACCGGTGAAAAGCACTAAGAAAGGCGACCTGATCGATCCCTTA 1963
Oy      1849 TATGTTGACGCTACAGTTGAGCTTAAAAACGGTATTTGATTTCTCATAGGATCA 1908
Db      1964 CAATCTCAATATATCTGTAGAAAGTCAAAAAACGGTATTTAATCATACCTCATTTAGACA 2023
Oy      1909 TTACCATATATATTAATTTGCTGTTGATGATCACATPACAAAGCTCCAAATGGGCTA 1968
Db      2024 TTACCATATACATCAAAATTTGAGTGTGTAAGCAAGGCCCTTATGAGGACCTTAAGGGTA 2083
Oy      1969 TACCTTGAAGATTTGTTTGGCAGCATTAAGTACTACGTAGAACACCTGACGAACGTCC 2028
Db      2084 TACTCTGAGGATCTTTTGGGAGCTGTCAAGTACTATGTGCAACATCCAAAGCAACGTCC 2143
Oy      2029 ACATTTCTATGATGATGGGCGCATGCAAGTGCATGTGTTAGCAAGAAAACACACAG 2088
Db      2144 GCATTCAGATATATGTTTGGTAAACGTAGCGACCATGTTCAAAGAAAACAAAATGCTCA 2203
Oy      2089 TGAAGATCCAAATAGAACTTCAAGCGGATGAGAG----- 2125
Db      2204 AGCTGATACCAATCAAAACGAAAAACCAAGCGAGAGAAACCTCAGACAGAAAAACCTGA 2263
Oy      2126 ----- 2125
Db      2264 GGAAGAAACCCCTCGAAGAGAAACCAAAAGCAGAAACACAGAGTCTCCAAAACCAAC 2323
Oy      2126 -----CGAGTAGAGAAACACCTGCTGAGCCAGAGTCCCTCAAGTAGAGACTGAAA 2178
Db      2324 AGAGAACACAGAAAGAAATCACACAGAGAAATCAAGAAACCTCAGCTCGAGACTGAAA 2383
Oy      2179 AGTAGAACCCCAACCAAGAAAGCAAGATTTTGTTCGAAAGTAAGGATTCAGTCT 2238
Db      2384 GGTGTAAGAAAAAAGTGAAGAGGCTGAAGATTTCCTGAAAAATCCAGAGATCCAAATTAT 2443
Oy      2239 GAAAGCCCAATGCAACAGAAACCTAGCTGTTTGAAGATATATTTGACTCTTCAATATAT 2298
Db      2444 CAAGTCCATGCGCAAGAGAGCTCTCAGAGATTAATAATAATTACTATTTGGACCCCA 2503
Oy      2299 GGATTAACAATAGTATCATGCGAAGAGCAAGAAAATTAATCTTGCTGTTTAAAGAGATTA 2358
Db      2504 GGACAAACAATACTATTATGCGAAGAGCTGAAAAAATAATTGCTTATTAAGAGAGATTA 2563

```

```

RESULT 9
ABK15103
ID ABK15103 standard; DNA; 2647 BP.
XX
XX ABK15103;
AC
XX
XX
DT 08-MAY-2002 (first entry)
XX
DE DNA encoding Streptococcus pneumoniae BVH-11.
XX
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX pneumonia; streptococcal bacterial infection; gene; ds.
XX
XX Streptococcus pneumoniae.
OS
XX
FH Key Location/Qualifiers
FT CDS 45..2567
FT /*tag= a
FT /product= "BVH-11"
FT /note= "The gene is flanked by sequences from the
FT vector SP64, no information on which is
FT given in the specification"
XX
XX
XX MO200198334-A2.
XX
XX
XX PD 27-DEC-2001.
XX
XX 19-JUN-2001; 2001WO-CA00908.
XX
XX 20-JUN-2000; 2000US-212683P.
XX
XX (SHIR-) SHIRE BIOCHEM INC.
XX
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX
XX WPI; 2002-122272/16.
XX
XX DR P-PSDB; AAU75933.
XX
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX epitope-bearing polypeptides, useful as vaccine components for treating
XX or preventing streptococcal infections such as otitis media,
XX meningitis, and bacteraemia
XX
XX
XX Disclosure; Fig 4; 113pp; English.
XX
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
XX identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX comprising (I) is useful for therapeutic or prophylactic treatment of
XX meningitis, otitis media, bacteraemia or pneumonia infection in an
XX individual susceptible to these disorders. (II) is also useful for
XX therapeutic or prophylactic treatment of any streptococcal bacterial
XX infection (e.g., caused by Streptococcus pneumoniae, group A
XX Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
XX as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or
XX Streptococcus aureus) in an individual susceptible to the infection.
XX A polynucleotide (III) encoding (I) is useful in DNA immunisation
XX techniques. The Streptococcus polypeptides are useful in a diagnostic
XX test for S. pneumoniae infection. (III) is useful for designing DNA
XX probes for use in detecting the presence of Streptococcus in a biological
XX sample suspected of containing the bacteria. The DNA probes may also be
XX used for detecting circulating S. pneumonia nucleic acid in a sample for
XX diagnosing streptococcal infections. This sequence encodes the
XX Streptococcus pneumoniae protein BVH-11, used to create the antigenic
XX peptides described in the method of the invention.
XX
XX
XX Sequence 2647 BP; 934 A; 538 C; 556 G; 619 T; 0 other;

```

```

Query Match 57.5%; Score 1374.2; DB 24; Length 2647;
Best Local Similarity 73.9%; Pred. No. 0;
Matches 1819; Conservative 0; Mismatches 539; Indels 102; Gaps 2;

```

QY 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAAAGGTTAAGAAAAATATCGTGTCTTA 60
 Db 104 TCGTTATGAACTAGAGTTTGCATCAAGCTCAAACTGTAAAAAGAAAAATATCGTGTCTTA 163
 QY 61 TATAGATGAAAAAACAAGCAAGCAAAAGGAGAAATTTGACTCTCGTATGAGTTAGCA 120
 Db 164 TATAGATGAAAAAACAAGCAAGCAAAAGGAGAAATTTGACTCTCGTATGAGTTAGCA 223
 QY 121 GGGTAGAGAAATCAATGCTGAGCAAAATCGTCAATCAAGATATACAGAACCAAGGCTATGTC 180
 Db 224 GGGTAGAGAAATCAATGCTGAGCAAAATCGTCAATCAAGATATACAGAACCAAGGCTATGTC 283
 QY 181 TTCAATGAGCAAGCACTATCATTTATACAAATGCTGAGTTGCTCTTATGAGCTATCATG 240
 Db 284 CTCTCATGAGAACATTATCATTTATGAGCAAGGCTCTTATGAGCTATCATG 343
 QY 241 TGAAGAAATTAATCTGATGAAAAAGTCAAACTATTAAGCTTAAAGATGAGGATTTGTTAAGA 300
 Db 344 TGAAGAAATTAATCTGATGAAAAAGTCAAACTATTAAGCTTAAAGATGAGGATTTGTTAAGA 403
 QY 301 GGTCAAGGAGTGAATGTTATCAAGGTAGATGAAAAATATGTTTACTTAAAGATGTC 360
 Db 404 AATCAAGGAGTGAATGTTATGTTATTAAGTTAAAGGTTAACTATGTTTACTTAAAGATGTC 463
 QY 361 TGGCCAGCGAGTAAAGTCCGTCAAAAGAGAAATCAATCAAGCAAAAGAGCATAG 420
 Db 464 AGCTCATGCGAGTAAAGTCCGTCAAAAGAGAAATCAATCAAGCAAAAGAGCATAG 523
 QY 421 TCAAACTGTTGAAGTGAAGTCTCAAGAAACGATGCTGCTGCTTGGACGTTGCGA 480
 Db 524 TCAGATCGTTGAAGGAGGAGCTTCAAGAAACGATGCTGCTGCTTGGACGTTGCGA 583
 QY 481 AGGAGCTTACTACAGATGATGTTATATCTTTAATGCTTCTGATATCATGAGATGTC 540
 Db 584 GGGAGCTTACACACAGATGATGTTATATCTTTAATGCTTCTGATATCATGAGATGTC 643
 QY 541 TGGTAGGCTTATATGCTTCTCATGAGATCATTAACATTAATCATTCCTTAAGATGAT 600
 Db 644 GGGTAGGCTTATATGCTTCTCATGAGATCATTAACATTAATCATTCCTTAAGATGAT 703
 QY 601 ATCAGTAGCAGATGTTGGTCTGTCAGAAAGCTTCTATCTGTCGAGAAATCTGTCAA 660
 Db 704 ATCAGTAGCAGATGTTGGTCTGTCAGAAAGCTTCTATCTGTCGAGAAATCTGTCAA 763
 QY 661 TTCAAGAACTTATCGCCGCAAAATGCGATTAACATTAAGAAACAACTGGTACCTTC 720
 Db 764 TTCAAGAACTTATCGCCGCAAAATGCGATTAACATTAAGAAACAACTGGTACCTTC 823
 QY 721 TGTAAAGCAATCCAGAACTCAAAATCTTAACAACAGCAACAGCAACATTAACAGTCA 780
 Db 824 TGTAAAGCAATCCAGAACTCAAAATCTTAACAACAGCAACAGCAACATTAACAGTCA 883
 QY 781 AGCAAGTCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
 Db 884 AGCAAGTCAAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 943
 QY 841 TCAAGCACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 Db 944 TCAAGCACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1003
 QY 901 AGCTGAGGTTGTCAGTCCAGAGAGATCATTAACATTAACATTAACATTAACATTAAC 960
 Db 1004 AGCTGAGGTTGTCAGTCCAGAGAGATCATTAACATTAACATTAACATTAACATTAAC 1063
 QY 961 GTCTGATTTGAAGAAAGATGCTGATGATGATGATGATGATGATGATGATGATGATG 1020
 Db 1064 GTCTGATTTGAAGAAAGATGCTGATGATGATGATGATGATGATGATGATGATGATG 1123
 QY 1021 GGTACCAAGTTCAAGCCAGAAACCAAGTCCAAACAGGATCTCGGAACCTATGTCAG 1080
 Db 1124 GGTACCAAGTTCAAGCCAGAAACCAAGTCCAAACAGGATCTCGGAACCTATGTCAG 1183

QY 1081 CCCGCAAGCTGCAACCAATCTTAAATAGACTCAA-----ATTCCTTTGGT 1128
 Db 1184 TCCGCAAGCTGCAACCAATCTTAAATAGACTCAA-----ATTCCTTTGGT 1243
 QY 1129 TAGTCAGCTGTCAGAAAGTTGGGAGAGATGATGATGATGATGATGATGATGATGATG 1188
 Db 1244 CAAGAAAGCTGTCAGAAAGTTGGGAGAGATGATGATGATGATGATGATGATGATGATG 1303
 QY 1189 TTATGCTTTTGGGAAAGTTTACATCTGAAAGCTTTAAAAATCTTGAAGCAATGATATC 1248
 Db 1304 TTATGCTTTTGGGAAAGTTTACATCTGAAAGCTTTAAAAATCTTGAAGCAATGATATC 1363
 QY 1249 AAAACAAGAGAGTGTTCACACATTAAGTCTTAAAGAAAGAAATGTTGCTCCGTCGA 1308
 Db 1364 CAAGCAAGAGAGTGTTCACACATTAAGTCTTAAAGAAAGAAATGTTGCTCCGTCGA 1423
 QY 1309 CCAGAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1368
 Db 1424 TCAGAAATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1483
 QY 1369 AAATGAGGTCGTAAATTCGATTTCCAGAGCTTGAACAAATTAATGAAAGCTTGAATGA 1428
 Db 1484 TAATGAGGTCGTAAATTCGATTTCCAGAGCTTGAACAAATTAATGAAAGCTTGAATGA 1543
 QY 1429 TGAATGCACTAATAAGAAAAATGCTGATGATGATGATGATGATGATGATGATGATGAT 1488
 Db 1544 TGTCTCAAGTATTAAGTCAAGTATGATGATGATGATGATGATGATGATGATGATGAT 1603
 QY 1489 CCATCCAGAGCACTTGGCAACCAAAATTCCTAAATGATGATGATGATGATGATGATGAT 1548
 Db 1604 TCATCCAGAGCACTTGGCAACCAAAATTCCTAAATGATGATGATGATGATGATGATGAT 1663
 QY 1549 TATTCCTCAATTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1608
 Db 1664 AGTACCAAGTGTGAGGCAAGTATCAACAGAAACGCTTATTAATCTTGAATCCTGCTGA 1723
 QY 1609 TATATCAAGTATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1668
 Db 1724 TATATCAAGTATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1783
 QY 1669 TGAAGAAATGAGCTTGTCTGATTAAGAAAAAGTTGCAAGCTCAAGCTTATCTTAAAGAAA 1728
 Db 1784 TGAAGAAATGAGCTTGTCTGATTAAGAAAAAGTTGCAAGCTCAAGCTTATCTTAAAGAAA 1843
 QY 1729 AGTATCTTACTCTCATCTCCAGAGCAAGTGTAAACCAATCCAACTGAGATGATGTC 1788
 Db 1844 AGTATCTTACTCTCATCTCCAGAGCAAGTGTAAACCAATCCAACTGAGATGATGTC 1903
 QY 1789 AGCAGCTATTTAACAATCGTGTGAAGGGGAAAAAGCAATTCACACTGTTCCACTTCA 1848
 Db 1904 AGCAGCTATTTAACAATCGTGTGAAGGGGAAAAAGCAATTCACACTGTTCCACTTCA 1963
 QY 1849 TATGTTGAGCATACAGTTGAGTTTAAAGCGTAAATTTGATTAATCTCTCATTAAGATCA 1908
 Db 1964 TATGTTGAGCATACAGTTGAGTTTAAAGCGTAAATTTGATTAATCTCTCATTAAGATCA 2023
 QY 1909 TTAACATTAATTAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGAT 1968
 Db 2024 TTAACATTAATTAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGAT 2083
 QY 1969 TACCTTGAAGATTTGTTGCGAGATTAAGTACTAGTGAAGCAACCTGACGAAAGTCC 2028
 Db 2084 TACCTTGAAGATTTGTTGCGAGATTAAGTACTAGTGAAGCAACCTGACGAAAGTCC 2143
 QY 2029 ACATTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2088
 Db 2144 ACATTTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2203
 QY 2089 TGAAGATCAATTAAGAACTTCAAGCGGATGAAG----- 2125
 Db 2204 TGAAGATCAATTAAGAACTTCAAGCGGATGAAG----- 2263
 QY 2126 ----- 2125

QY 1018 TTGGGTACAGATTCAAGGCCAGAACCAAGTCCAGAACCGACTCCGAAACCTAGTCC 1077
 DB 1038 TTGGGTACAGATTCAAGGCCAGAACCAAGTCCAGAACCGACTCCGAAACCTAGTCC 1097
 QY 1078 AAGCCCGCACTGCACTCAAAATCTTAAATAGACTAA-----ATTCTTCTTT 1125
 DB 1098 AAGTCGCAACCTGCACTCAAAATCTTAAATAGACTAA-----ATTCTTCTTT 1157
 QY 1126 GGTAGTCACTGCTGCAAGAAAGTTGGGAGGATATGATTGCAAGAAAGGACATCTC 1185
 DB 1158 GGTCAAGAAAGCTGTTGCAAGAAAGTTGGGAGGATATGATTGCAAGAAAGGACATCTC 1217
 QY 1186 TCGTATATGCTTTGGCAAGAAATTTACATCTGAACTGTTAAATCTTGAAGCAAGTT 1245
 DB 1218 TCGTATATGCTTCCAGCCAGAGATCTTACAGAAACAGACAGGACTGATTAAGCAACT 1277
 QY 1246 ATCAAAACAGAGAGTGTTCACACCTTTAACTGTTAAAGAAAGAAATGTTGCTCTG 1305
 DB 1278 GGCACAGCAGAAAGTTATCTCATAGCTAGAGCTAAGAAACTGACCTCCCATCTAG 1337
 QY 1306 TGACCAAGAAATTTATGATTAAGCATATATCTGTTAAGAGGCTCATTAAGCCTGTT 1365
 DB 1338 TGATGAGAAATTTTACAAATTAAGCTTATGACTTACAGAAAGATTCACCAAGATTACT 1397
 QY 1366 TGNAAATTAAGGTCGTATTTCTGATTTCCAGCCTTAGACAAATTTAGAAAGCCTTGA 1425
 DB 1398 TGATATTAAGGTCGACAAAGTGTATTTGAGGCTTTGATTAACCTGTTGAAGCACTCA 1457
 QY 1426 TGATGATGACTAATTAAGAAATTTGATGATGATTTATTTGCGACTTCTAGCACCAAT 1485
 DB 1458 GATATGTCGCAAGTGAATTAAGTCAAGTATGATGATGATTTCTGCTTCTAGCTCGAT 1517
 QY 1486 TACCATCCAGAGCAGCTTGGCAACCAAAATCTCAAAATGATGATGATGATGATGAT 1545
 DB 1518 TCGTATCCAGAACTGTTGAAACCAATTAAGGCAATTAACCTGATGATGATGATGAT 1577
 QY 1546 TCGTATGCTCAATTAAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1605
 DB 1578 TCAAGTAGCAGAGTTGGCAGGAGCAAGTACCAACAGAAAGCGGTTATCTTTGATCTCG 1637
 QY 1606 TGATATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1665
 DB 1638 TGATATTAACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1697
 QY 1666 GATTGGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1725
 DB 1698 GATTAAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1757
 QY 1726 AAAAGGATCTTACCTCCATCTCCAGACGCAAGTGTAAAGCAATCCAACTGAGATAG 1785
 DB 1758 GAAAGGTTGACCCCTCTCTTCCAGACGCAATCCAACTGAGATGAGATGAGATGAG 1817
 QY 1786 TGCAGCAGCTATTTTACATCGTGTGAAGGAGGAAAAAGCAATTCACCTGTCGACTTC 1845
 DB 1818 AGCAGAGGCTATCTCAACCGCGTGAAGCAGCTAAGAGGAGGCTTGTATGCTATGCTC 1877
 QY 1846 AATATGCTGAGCATACAGTGTGAAGTAAAAAGGTAATTTGATTAATCTCTCATTAAGA 1905
 DB 1878 TTACATCTTCAATATACGTGAAGTCAAAAAAGGTAATTTAATCATCTCATTAATTA 1937
 QY 1906 TCATTAACATTAATTAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGAT 1965
 DB 1938 CCATTAACATTAATTAATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1997
 QY 1966 CTATACCTTGAAGATTTGTTGCGAGATTAAGTACTGATGAAGACACCTGACGAGAG 2025
 DB 1998 GTATATCTTGAAGATCTTTTGGCGAGCTGTCAAGTACTGATGATGATGATGATGATGAT 2057
 QY 2026 TCCACATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2085
 DB 2058 TCCGATTTAGATTAAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 2117
 QY 2086 CAGTGAAGATCCAAATTAAGAACTTCAAGCGGATGA----- 2121

DB 2118 TCAGCTGATATCAATCAAAACGAGAAACCAAGCAGAGAGAAACCTCAGACAGAAAAACC 2177
 QY 2122 ----- 2121
 DB 2178 TGAGGAGAAACCCCTCGAGAGAGAAACCGCAAGCAGAGAAACCGAGATCTCAAAAACC 2237
 QY 2122 ---AGAGCCAGTGAAGAGAAACCACTGAGCAGAGATCTCTCAAGTGAAGTGAAGTGA 2178
 DB 2238 AACAGAGAGAACCGAGAGATTCACAGAGAGATTCAGAGAAACCTCAGAGTGAAGTGAAGT 2297
 QY 2179 AGTAGAAGCCCACTCAAG 2238
 DB 2298 GGTGAAGAGAAACCTGAG 2357
 QY 2239 GAAAGCCATGCAAG 2298
 DB 2358 CAAGTCCATGCAAG 2417
 QY 2299 GGATACATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2358
 DB 2418 GGCACAACTATTAATTAAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2477
 RESULT 11
 AAA05417
 ID AAA05417 strand: DNA; 2481 BP.
 XX
 AC AAA05417;
 XX
 DT 24-MAY-2000 (first entry)
 XX
 DE Streptococcus pneumoniae nucleotide sequence ID311.
 XX
 KW Streptococcus pneumoniae; vaccine; screening; protein antigen;
 KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;
 KW pneumococcal disease; ds.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200006737-A2.
 XX
 PD 10-FEB-2000.
 XX
 PF 27-JUL-1999; 99MO-GB02451.
 XX
 PR 27-JUL-1998; 98GB-0016337.
 PR 19-MAR-1999; 99US-0125164.
 XX
 PA (MIGR-) MICROBIAL TECHNIQS LTD.
 PI
 PI Gilbert CFG, Hansbro PW,
 DR WPI: 2000-195300/17.
 DR P-PSDB; AAY81662.
 XX
 PT New Streptococcal protein, useful as a vaccine, for diagnosis of
 PT pneumococcal diseases and for screening agents capable of antagonizing
 PT or inhibiting expression of the protein
 XX
 PS Claim 2; Page 99; 108pp; English.
 CC AAY81501 to AAY81679 represent specifically claimed protein sequences
 CC isolated from Streptococcus pneumoniae. AAA05407 to AAA05590 represent
 CC specifically claimed nucleotide sequences isolated from S. pneumoniae.
 CC The sequences have antibacterial and antiinflammatory properties.
 CC The protein sequences, and fragments of them, are useful as immunogens
 CC and/or antigens. The nucleotide sequences can be used in vaccines and in
 CC diagnostic assays. The proteins and nucleotides can be useful for the
 CC detection and diagnosis of S. pneumoniae. The protein sequences are also
 CC useful for screening an agent capable of antagonizing, inhibiting or
 CC interfering with the function or expression of the proteins in which the
 CC agent is useful for treatment or prophylaxis of S. pneumoniae infection

CC and meningitis. AAA05591 to AAA05614 represent primers used in the
XX exemplification of the present invention.

Sequence 2481 BP; 839 A; 514 C; 538 G; 590 T; 0 other;

Query Match 42.0%; Score 1003.8; DB 21; Length 2481;
Best Local Similarity 66.0%; Pred. No. 4.8e-243;
Matches 1623; Conservative 0; Mismatches 693; Indels 144; Gaps 6;

```
QY 1 TTCTTACGAGTTGGAGCTGTATCAGCTTGAACGGTTAAGGAAA---TAACTGTTTC 57
DB 60 TTCTTATAGCTGTGGTGTCTACACCAAGCTGTCAGAGATAAGAAAGTCTTAATCGAGTTGC 119
QY 58 CTATATAGATGAAAAAACAAGCGACCAAAAACGAGAAATTTGACTCTGATGAGGTAG 117
DB 120 TTATATAGATGTTGATCAGCTGTGTCMAAAGGACAAAACTTGACACCGAGTGAAGTCA 179
QY 118 CAAGCGTGAAGGAATCAATGCTGAGCAAACTGTCATCAAGATTAAGACCAAGGCTATGT 177
DB 180 TAAAGAGGAGGGGATCAAGCGCGAACAATGTCATCAAGATTACGGATCAAGGTTATGT 239
QY 178 CACTTACATGGCGACCACTATCATTTATCAATGTTAAGGTTCTTATGAGCTATCAT 237
DB 240 GACCTTCTATGAGACCAATTATCATTAATGCAAGGTCCTTTATGATGCATCAT 299
QY 238 CAGTGAAGAATTACTCATGAAGATCCAAACTATTAAGCTTAAAGATGAAGATTTGTTAA 297
DB 300 CAGTGAAGAGCTCTCATGAAGATCCGAATTTATCAAGTTGAAGATTCAGACATTTGTCA 359
QY 298 TGAGGTCAAGGGTGAATATGTTATCAAGGTAGATGAAAAATACTATGTTTACTTTAAGA 357
DB 360 TGAATTAAGAGGTGGTATGTTATCATCAAGGTAGACGAAAACTATGTTTACTTTAAGGA 419
QY 358 TGCTGCCCCGCGGATTAAGTCCGTTACAAAAGAGAAATCAATCGACAAAAACAAGAGA 417
DB 420 TGCACCTACTGCGGATTAATATTCGACAAAAGAAAGATTTAACTGCAAGACAGAAAG 479
QY 418 TAGTCAACATCGTGAAGGTGAACCTCAAGAAACGATGCTGTTGCTTGGCGACGTTTC 477
DB 480 CAGTCAATATCAACGGGCTCAGAGCT-----AACATCATGCAAGTAGTGCAGCCAGAGC 533
QY 478 GCAGAGCGCTTACTACAGATGATGTTATATCTTTAATGCTTGTGATATCATAGAGA 537
DB 534 CCAAGAGCGCTTACTACAGATGATGTTATCTTTCAATGCAATGCAATCATTTAGAGA 593
QY 538 TACTGTGATGCTTATATGTTCTTCATGAGATCATTAACCTTACATTTCTTAAGAAAGA 597
DB 594 CACGGGTATGCTTATATGTTCTTCACGCGACCAATTAACATTCATTTCTTAAGAAAGA 653
QY 598 GTTATCAGCTAGAGTTGGTGTGTCGCAAGAGCTTCTCATGTCGCGAGAAATCTGTC 657
DB 654 GTTATCAGCTAGAGTTGCTGTGTCGCAAGAGCTTATGGAATGG----- 698
QY 658 AAATTCAGAAACCTATCGCCGACAAATAGCGATTAACCTTCAGAAACAACTGGGTACC 717
DB 699 -GAACAGAGGATCTGCTCTTCTTCAAGTTCTAGTTATATATGCAATCAGCTCAACAA 757
QY 718 TTCTTGAAGCAATCCAGAACTCAAAATTAACAACAAGCAACAACAGACACTTAACAG 777
DB 758 GATTGTCAAGAAACCAACATCT-----GACTGTCACTCCAACTTA 797
QY 778 TCAACAGATCAAGATTAAGATGATGATCTCTTGAAGAGCTCTCAAACTGCGCTTT 837
DB 798 TCATCAAAATCAAGGGGAAACAATTTCAAGCTTTTACGTGAATGTATGCTTAAACCTTT 857
QY 838 GAGTCACAGACATGATGATGATGAGCTTGTCTTTGATTCAGACCAATCAACAGTGC 897
DB 858 ATCAGAGAGCATGATGATGATGAGCTTGTATTTTTCAGCCAGGCAAAATCAACAGTGC 917
QY 898 AACAGCTAGAGGTGTCAGTGCACACGAGATCATTAACACTTACCTTACTCTCA 957
DB 918 AACCCCGAGAGGTGATGCTGCTCCATGATGATTAACCATTAACCTTATTCCTTAAGACA 977
```

```
QY 958 AATGCTGAATTGGAGAACAAGATCGCTGATATTATCCCTTCGTTATGTTCAAAACA 1017
DB 978 AATGCTGAATTGGAAAAAGAAATGCTGATATTATCCCTTCGTTATGTTCAAAACA 1037
QY 1018 TTGGGTACAGATTTCAAGGCCAGACAAACCAAGTCCCAACCCGACTCCGAACTAGTCC 1077
DB 1038 TTGGGTACAGATTTCAAGGCCAGACAAACCAAGTCCCAACCCGACTCCGAACTAGTCC 1097
QY 1078 AGGCGCGCAACCTCAGCAAAATCTTAATAATGACTCA-----ATTCCTCTT 1125
DB 1098 AAGTCCGCAACCTCAGCAAAATCTCACCAGCTCCAGCAATCAATTGATGAAATTT 1157
QY 1126 GGTATGACGTGTGTACAAAAGTTGGGAGATATGTAATTCAGAAAAAGGCACTTC 1185
DB 1158 GGTCAAGAAAGCTGTGCAAAAGTAGCGATGTTATGCTTTGAGGAGAAATGAGAGTTTC 1217
QY 1186 TCGTTATGCTCTTGGGAAAGTTTACATCTGAACCTGTTAAAAATCTTGAAGCAAGTT 1245
DB 1218 TCGTTATATCCACCCAAAGATCTTTGAGCAAAAACAGACAGGCAATGATTAACAACT 1277
QY 1246 ATCAAAACAAGAGGTGTTTCAACACTTAACTGCTTAAAAAAGAAATGTGCTCTCG 1305
DB 1278 GCGCAAGAGGAAAGTTTATCTCAATAGCTAGAGCTAAGAAAACTGACCTCCATTTAG 1337
QY 1306 TGACCAAGAAATTTATGATTAAGCATTAATCTGTTAATGAGGCTCATTAAGCTTGT 1365
DB 1338 TGATCGAAGATTTACATATAAGCTTATGACTTATGCAAGATTCACCAAGATTTACT 1397
QY 1366 TGNAAATPAAGGCTGTAATTCGATTTCCAAAGCTTGAACAAATTTAGAACCTTGAA 1425
DB 1398 TGATTAATPAAGGTGACAAAGTTGAGGCTTTGGAATTAACCTGTTGGAACCACTCAA 1457
QY 1426 TGATGAATTCACATTAAGAAAAATGATAGATATTTATGTCATTCCTAGCAACAA 1485
DB 1458 GGATGTCCTCAAGTATTAAGCTAAGTAGAGATATTTGCTTCTTATGCTCGAT 1517
QY 1486 TACCATTCAGAGGAGCTTGGCAAAACCAATTCCTCAATGATGATTAAGTGAAGAGAGT 1545
DB 1518 TCGTATTCAGAAAGTTTAGAAAAACCAATGCGCAATTAACCTGATGATGAGAT 1577
QY 1546 TCGTATGCTCAATTAGCTGATTAAGTATTAACAGCTCAAGTGTATCAATTTTGAAGACA 1605
DB 1578 TCAAGTAGCCAAAGTTGCGAGGCAAGTACACAAACGAAAGCGGTTATATCTTGAATCTCG 1637
QY 1606 TGATATATACAGATGAAGAGATGATATGTAAGCTCATATGAGGCCATAGTACTG 1665
DB 1638 TGATATTAACAGATGATGAGGAGATGCTATGTAATCTCAATATGACCACTAGCACTG 1697
QY 1666 GATTGAAAAAGATAGCTTTCTGATTAAGAAAAAGTTGCAAGCTCACTATTAAGA 1725
DB 1698 GATTAAAAAGATAGTTTGTCTGAAGCTGAGAGAGCGGACCGCAAGCTTATGCTTAAGA 1757
QY 1726 AAAAGTATCTTACCTTCATCTCAGACGAGATGTTAAAGCAATCCAACTGAGATAG 1785
DB 1758 GAAAGTTTGAACCTCTCTTCGACAGACCAATCAGAGATTCAGGAATATCTGAGCAAAAG 1817
QY 1786 TGCAGACGCTTATTAACAATGCTGGAAGAGGAAAAAGAAATTCACCTGCTGCACTTCC 1845
DB 1818 AGCAGAAAGCTATTAACAACGCGTGAAGAGCTTAAGAGGTGCACCTGATGATGCC 1877
QY 1846 ATATATGTTGAGCATCAGTTGAGGTTAAAAACGTTAATTTGATTAATCTCTCAATAGA 1905
DB 1878 TTACATCTTCAATATATGATGATGAGTCAAAAAAGGATGTTATCACTCACTCATTTAGA 1937
QY 1906 TCATTAACATTAATTAATTTGCTTGTGTTGATGATCAACATTAACAAAGTCCAAATGG 1965
DB 1938 CCATTAACATTAACATTAATTTGATGTTGATGAGAGGCTTTATATAGGCACTTAAGGG 1997
QY 1966 CTATACCTGGAATAATTTGTTGGAGATTAATTAATCTAGCTAGTAACACCTGAGCAAG 2025
DB 1998 GTATACCTTTGAGGATCTTTTGGAGCTGTCAACTATCTAGTCAACATTCACCAAGGAG 2057
QY 2026 TCCACATTTCTAATGATGATGAGGCAATGCCAGTGAACATGTTTGAAGCAAGAAAGACA 2085
```

```

Db      2058 TCCGATTAGATTAATGTTTGGTAAAGCTACCGCATGTTCAAAAGAAACAAAATATG 2117
OY      2086 CAGTGAAGATCCAAATPAAGAACTTCAAGCGGATGA----- 2121
Db      2118 TCAAGGTATACCAATCAAAAGGAAAAACCAAGCGGAGAAACCTCAGACAGAAAAACC 2177
OY      2122 ----- 2121
Db      2178 TGAGGAGAAACCCCTCGAAGAGAAACCGCAAGCGAAGAACAGAGCTTCCAAACC 2237
OY      2122 ---AGAGCCATGAGAGAAACACCTGCTGAGCCAGAAAGTCCCAAGTGAAGCTGAAA 2178
Db      2238 AACAGGGAACGAGAAATTCACAGAGAAATCAGAAAGAACTCAGAGTGAAGCTAAAA 2297
OY      2179 AGTAGAAGCCCACTCAAAAGAGCAAGATTTTGGTGAAGAAAGTAAGGATTTAGTCT 2238
Db      2298 GGTGAGAGAAACCTGAGAGAGGCTGAGAAATTTACTTGAAGAAATCAGAGATCCAAATTAT 2357
OY      2239 GAAAGCCATGCAACAGAAACTCTAGCTGTTTACGAAATTAATTGACTCTTCAATTTAT 2298
Db      2358 CAAGTCCAAATGCAAGAGACTCTCAAGAGATTAAAAATTAATTACTATTGGCACCA 2417
OY      2299 GGATTAACATAGTATCATGCGAGAGACAGAAATTAATTGCGTTGTTAAAGAAAGTAA 2358
Db      2418 GGACAAACAATACTATTATGAGAGAGCTGAAAAAACTATTGGCTTTATTAAGAGAGTAA 2477

```

RESULT 12

ABX06885
ID ABX06885 standard; DNA; 2457 BP.

ABX06885;
XX

11-FEB-2003 (first entry)
XX

S. pneumoniae type 4 strain coding region #1173.
XX

Gene: ds; bacterial meningitis; pneumonia; sepsis; otitis media;
KM ear infection; antiinflammatory; antibacterial; immunostimulant;
KM auditory; respiratory; gene therapy; vaccine.
XX

Streptococcus pneumoniae type 4 strain.
OS

MO20027021-A2.
PN

03-OCT-2002.
XX

27-MAR-2002; 2002WO-1B02163.
PF

27-MAR-2001; 2001GB-0007658.
PR

(CHIR-) CHIRON SPA.
PA

(GENO-) INST GENOMIC RES.
XX

Masignani V, Tettelin H, Fraser C;
PI

MP1: 2003-040579/03.
XX

P-PSDB; AB001597.
DR

New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT useful as medicaments for treating or preventing a disease or infection
PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
PT or ear infection -
XX

Claim 6; SEQ ID NO 2345; 56bp; English.
PS

The invention relates to a protein comprising or having at least 50%
CC identity to any of the 2469 amino acid sequences, identified in the
CC specification (available on a computer readable format), or its fragment,
CC expressed from 2469 of 2489 identified DNA coding regions from the
CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
CC ABS56454. Also included are an antibody which binds one of the

CC proteins, treating a patient by administering the protein, DNA or
CC antibody (in a composition), a kit comprising first and second primers,
CC which are the nucleic acid cited above or fragments between nucleotides
CC 8-100 of a sequence not defined in the specification, for amplifying a
CC target sequence contained within a Streptococcus nucleic acid sequence,
CC where the first primer is substantially complementary to the target
CC sequence and the second primer is substantially complementary to the
CC complement of the target sequence, and where the parts of the primers
CC having substantial complementarity define the termini of the target
CC sequence to be amplified, assay comprising contacting a test compound
CC with the protein, and determining whether the test compound binds to the
CC protein and a Streptococcus pneumoniae bacterium, where one or more
CC genes encoding the proteins has been rendered inactive. The proteins,
CC nucleic acid molecules, antibody and compositions are useful as
CC medicaments for treating or preventing a disease or infection due to
CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
CC sepsis, otitis media or ear infection. They are also useful in developing
CC vaccines, diagnostics and antibiotics. The methods are useful for
CC identifying immunodominant proteins. The present sequence is one of
CC the 2489 identified coding region from the genomic sequence.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIP0 at
CC ftp.wipo.int/pub/published_pct_sequences.
XX

Sequence 2457 BP; 836 A; 497 C; 531 G; 593 T; 0 other;

Query Match 41.6%; Score 993; DB 25; Length 2457;

Best Local Similarity 65.8%; Pred. No. 2.6e-240;

Matches 1611; Conservative 0; Mismatches 696; Indels 141; Gaps 6;

```

OY      1 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAGACGGTTAAGGAAA---TAATCGTGTTC 57
Db      60 TTCTTATGAGCTTGAAGCTTACCAAGCTGTCAGGATTAAGAAAGCTTAAATGAGTTC 119
OY      58 CTATATAGATGAAAAACAAGCGCAAAACGAGAAATTTGACTCTGATGAGTTAG 117
Db      120 TTATATAGATGATGATCAGGCTGTCGCAAAAGCAAAAACCTGACACCAAGATGAGTCA 179
OY      118 CAAGGTGAAGAAATCAATGCTGAGCAAAATGCTCATCAAGATTAACAGCAAGCTATGT 177
Db      180 TAAAGGAGAGGGGATCAACGCCGAACAAATGTTATCAAGATTACGAGTCAAGGTTATGT 239
OY      178 CACTTCACATGGCGACACTATATTATTAACATGATGATGAGTTCTTATGAGCTATCAT 237
Db      240 GACCTCTATGAGAGCAATTTATCTATTATATGCAAGGTTCTTATGATGACATCAT 299
OY      238 CAGTGAAGATTTACTCATGAAAGATCCAAACTATTAAGCTTAAAGATGAGATATTGTTAA 297
Db      300 CAGTGAAGAGCTCCCATGAAAGATCCGAATTTATGAGATGAGATGAGATGATGATCA 359
OY      298 TGAGGTCAAGGGTGAATGTTATCAAGGTAGAGTAAATTAATCTATGTTTACCTTAAGA 357
Db      360 TGAATTCAGAGGTGATGTTATGATTAAGTTAAACGGTAAATCTATGTTTAAAGGA 419
OY      358 TGCTGCCACCGCGATTAACGTCCTTACAAAGAGAAATTCATGACAAAAACAGAGCA 417
Db      420 TGACGCTATGGGATTAATTTGCGACAAAGAAAGATTAACGTCGAGAGCAGAAACG 479
OY      418 TAGTCAACATGCTGTAAGGTGAACTCCCAAGAAACGATGCTGTTGCTGGCAGCTTC 477
Db      480 CAGTATATATAT-----AACTCAAGACAGATATATGTTGCTGACGACGAGAGC 530
OY      478 GCAAGAGCTTATCTACAGATGATGTTATATCTTTATATGCTTGTGATATCAGAGA 537
Db      531 CCAAGAGCTTATACACGAGATGATGAGGTATATCTTCAATGATCTGATATCATTTAGA 590
OY      538 TACTGATATGCTTATATGTTCTTCAAGAGATATATACATTTATCTTAAATAAGA 597
Db      551 CACGGGTATGCTTATATGTTCTTCAAGAGATATATACATTTATCTTAAATAAGA 650
OY      598 GTTATACACTGACGAGTGGTGTGACAGAGCCTTCTATCTGCTGACAGAAATCTGTC 657

```

Db 651 GTTATCAGTAGCGAGTTAGCTGCTGCAGAACCTATTGGAAATGG----- 695
Qy 658 AATTCAGAACCTATTGCGCGACAAATATGCGATTAACCTTCAAGAACCAATGGGTACC 717
Db 656 -GAGCAGAGGATCTGCTCTTCTTCAAGTTCTAGTTATATCAAAATCCAGCTCAACCA 754
Qy 718 TTCTGTAAAGCAATCGAGAACTACAAATTAACAACAAGCAACAGCAACACTAACAG 777
Db 755 GATTGTGAGAGAACCAATCT-----GACGTGCTCCCACTTA 794
Qy 778 TCAAGCAATGCAAAATATGATGATGATGCTCTTGAACAGCTCTACAAATCGCTTT 837
Db 795 TCATCAAAATCAAGGGGAAACATTTCAAGCTTTTACGTGAATGTATGCTAAACCTT 854
Qy 838 GAGTCACGACATGATAGATCTGATGAGCTGCTTGTATGATCCAGCAAAATCAGAATCG 897
Db 855 ATCAAGACCCATGTGGAATCTGATGCTTATTTTTCAGCCAGGCAAAATACAAAGTCG 914
Qy 898 AACAGCTAGAGGTGTTCAGTGCACACGAGATCATTAACAATTATCCCTTACTCTCA 957
Db 915 AACCCGAGAGGTGATGCTGCTCCATGTAACCATTAACAATTATCCCTTATGAGACA 974
Qy 958 AATGCTGATTTGGAAGACGATGCTGCTGATATATCCCTTCTGTTATGTTCAACCA 1017
Db 975 AATGCTGATTTGGAAGACGATGCTGCTGATATATCCCTTCTGTTATGTTCAACCA 1034
Qy 1018 TTGGGTACAGATTCAGAGCCAGAACCAACAGTCACAAAGGCTCCGGAACCTAGTCC 1077
Db 1035 TTGGGTACAGATTCAGAGCCAGAACCAACAGTCACAAAGGCTCCGGAACCTAGTCC 1094
Qy 1078 AGGCCCGCAACCTGCACCAAAATCTTAAATAGACTCAAAATCTTCTTGGTTAGTACGT 1137
Db 1095 AAGTCGCGAACCGAGCTCCAGCAATCCATTTGATGAGAA-----ATTGGTCAAAAGAAC 1148
Qy 1138 GGTACGAAAGTTGGGGAAGATATGATTCGAGAAAGGCGACTCTCTGTTATGCTT 1197
Db 1149 TGTTCGAAAGTAGGCGATGATGCTTGTGAGGAGATGAGATTTCTCGTTATATCC 1208
Qy 1158 TGCGAAGATTTACCATCTGAACGTTTAAATCTGGAAGCAAGTTATCAAAACAGA 1257
Db 1209 AGCCAAAGATCTTTACGAGAAACGACAGCAGCATGATACAAACCTGGCCAAAGGA 1268
Qy 1258 GAGTCTTCACACACTTTAACTGCTAATAAAGAAATGTTGCTCTGTCGACCAAGATT 1317
Db 1269 AAGTTATCTCTAAGCTAGAACTAAGAAACTGACCTCCATCTAGTATCGAAGATT 1328
Qy 1318 TTATATTAAGCATATATCTGTTAATGAGGCTCATAAAGCTGTTGTAATAAGG 1377
Db 1329 TTACATATAGGCTTATGACTTACTAGCAAGATTTGACCAAGATTTTATGATTAATAAGG 1388
Qy 1378 TCGTAATCTGATTTCCAAAGCCTTAGACAATTTATGAAGCCTGAATGATGATGCAG 1437
Db 1389 TCGACAGATGATTTTGAAGCTTTGATTAACCTGTTGAGACCATCAAGAGATGCTCAAG 1448
Qy 1438 TAATTAAGAAAAATTTGATGATTTATTTGCAATTCCTAGACCAATTAACCATCAGA 1497
Db 1449 TGATTAAGTCAGTTAGTGAAGATATCTTGCCTTTTAAAGTCCGATTCGTCACCGA 1508
Qy 1498 GCGACTTGGCAACCAAAATTTCTCAATGAGTATCTGAAGCAAGATTGCTATTGCTCA 1557
Db 1509 ACGTTTAGGAAACCAAAATGCGCAAAATTACTTACCTGATGATGATTTCAAGTGCACA 1568
Qy 1558 ATTACTGATTAAGTATCAACCTCGAGGTTAATTTTGTGTAACATGATATAAG 1617
Db 1569 GTTGGCAGGCAAGTACACAGAAAGACGTTATATCTTGTATCTCGATATATACAG 1628
Qy 1618 TGATTAAGGAGATGATATGTAACGCTCATATGGGCCATATCACTGATTTGAAAAAGA 1677
Db 1629 TGATTAAGGAGATGCTATGTAATCTCAATATGACCATACCATGATTTAAAAAGA 1688
Qy 1678 TAGCTTTCTGATTAAGAAAAAGTTGACGCTCAAGCTTATACTAAAGAAAAAGTATCT 1737
Db 1689 TAGTTTGTCTGAAGCTGAGAGAGCGGACGCGCTTATGCTAAAGAAAAAGTTTAC 1748

Qy 1738 ACCTCCATCTCCAGACCGAGATGTTAAAGCAAAATCCAATCGAGATAGTCAGACGTAT 1797
Db 1749 CCTCTCTTCGACACACCATCATGAGATTCAGAGAAATACAGAGCAAAAGAGCATAT 1808
Qy 1798 TTCAATGCTGTGAAGGGGAAAAAGAAATTCACCTGCTTGCATCTTATATGTTGA 1857
Db 1809 CTACAAACCGCTGAAACAGCTAAGAAAGTGCCCTGATGCTGTATGCTTCAATCTTCA 1868
Qy 1858 GCATACAGTTGAGTTAAAAAGGTAATTTGATATTCCTATAGAGATCTTACATTA 1917
Db 1869 ATATACGTAGAAAGTCAAAAAGGTAGTTATATATCTTATATGATCAATTA 1928
Qy 1918 TATTAATTTGCTGTTGTTGATGATCACATACAAAGCTCCAAATGCTTATACCTTGA 1977
Db 1929 CATCAATTTGAGTGCTTTGACGAAGGCTTTATAGAGCACTTAAGGGGTATACCTTGA 1988
Qy 1978 AGATTTGTTTCCGACGATTAATGATCTAGTAACACCTGACGACAGCTTCACTTCA 2037
Db 1989 GGATCTTTTGGCGACTGTCAAGTACTATGTCGAACATCCAAACGAAAGCTCCGATTCAGA 2048
Qy 2038 TGATGATGGGCAATGCGAGTGACATGCTTGAAGAAAGAACCAAGTGAAGATCC 2097
Db 2049 TAAATGTTTGTGTAACCTAGCGACATGTTCAAAAGAAACAAATGCTCAAGCTGATAC 2108
Qy 2098 AATTAAGACTTCAAGCGGATGA----- 2121
Db 2109 CAATCAAGCGAAACCAAGCGAGAAACCTCAGACGAAACCTGAGAAAGAAC 2168
Qy 2122 -----AGGCCAGT 2130
Db 2169 CCTCGAAGAGAAACCGCAACGAGAAACAGAGCTTCCAAACCAAGAGAAACC 2228
Qy 2131 AGAGAAACACCTGCTAGCGAGAGTCCCTCAGTGAAGCTGAATAAGTGAAGCCCA 2190
Db 2229 AGAAGAAATCACAGAGAAATGAGAAACCTCAGGTGAGCTGAAGTGAAGAAA 2288
Qy 2191 ACTCAAAAGACAGCAAGTTTGTGCGAAAGTAAAGGATCTTATGCTGAAGGCAATGC 2250
Db 2289 ACTGAGAGAGCTGAAGATTTACTTGAAGAAATCCAGATCCATTTATCAAGTCCATTC 2348
Qy 2251 AACGAAATCTAGCTGTTTACGAATTAATTGACTCTTCAATTAATGATTAACAATAG 2310
Db 2249 CAAAGACACTCTCAGAGATTAATAATTTCTATTGGCCACCCAGGACCAACAATAC 2408
Qy 2311 TATCATGCGAGAACGAGAAATTAATCTGCTGTTAAAGAGAGTAA 2358
Db 2409 TATTATGCGAGAACTGAAAACTATTGCTTTATTAAGAGAGATTA 2456

RESULT 13
ID AAA47605
AC AAA47605 standard; DNA, 2531 BP.
DT 20-OCT-2000 (first entry)
XX
DE Recombinant variant of Sp36 gene (Sp36B) of *S. pneumoniae*.
XX Streptococcus pneumoniae; infection; vaccine; coiled coil region;
XX histidine triad residue; Sp36; antibody; otitis media;
KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
KW meningitis; lobar pneumonia; ds.
XX Streptococcus pneumoniae.
OS
XX
XX Key Location/Qualifiers
XX 1. 2531
XX CDS /tag= a
XX /product= Sp36B polypeptide
XX
XX
XX PN W0200037105-A2.

XX 29-JUN-2000.
 PD 21-DEC-1999; 99WO-US0390.
 XX 21-DEC-1999; 98US-0113048.
 PR 21-DEC-1998; 98US-0113048.
 XX
 XX (MEDI-) MEDIMUNE INC.
 PA Johnson LS, Koenig S, Adamou JE;
 PI WPI, 2000-452129/39.
 DR P-PSDB; AAB01469.
 XX
 XX Vaccine useful for prophylaxis and treatment of pneumococcal infections
 PT as otitis media, nasopharyngeal and bronchial infections,
 PT comprises Streptococcus pneumoniae proteins
 XX
 PS Disclosure; Page 69-70; 70pp; English.
 XX
 CC Although a number of proteins have been suggested as being involved
 CC in the pathogenicity of Streptococcus pneumoniae, there still remains
 CC a need to identify polypeptides having epitopes in common from
 CC various strains of S. pneumoniae in order to utilize such
 CC polypeptides in vaccines to protect against a wide variety of
 CC S. pneumoniae. New vaccine compositions are described which comprise a
 CC Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino
 CC acids in length that comprise at least one histidine triad residue
 CC (HxxHxx) or a coiled-coil region, or an antibody directed against
 CC these features. The vaccine is useful in protecting against infection
 CC by Streptococcus pneumoniae. The vaccine composition comprising
 CC antibodies to its useful for passive immunization for treating
 CC pneumococcal infections which includes otitis media, nasopharyngeal
 CC and bronchial infections.
 XX
 SQ Sequence 2531 BP; 861 A; 508 C; 550 G; 609 T; 3 other;

Query Match 41.5%; Score 991; DB 21; Length 2531;
 Best Local Similarity 65.7%; Pred. No. 8.3e-240;
 Matches 1609; Conservative 1; Mismatches 697; Indels 141; Gaps 6;

QY 1 TTCTTACGAGTTGGAGCTGTATCAAGCTTGAAGGTTAAGAAAA--TATCGTGTTC 57
 DB 60 TTCCATATGAGCTTGGACGTTTACCAAGCTGTGTCAGATTAAGAAAGTCAATCGAGTTGC 119
 QY 58 CTATATAGATGAAAAAACAAGCAGCAAAACGAGAAATTGACTCTCGATGAGGTTTG 117
 DB 120 TTATATAGATGGTGTGATCAGGCTGTCTAAAGCAGAAAACTTGACACCGATGAAGTCA 179
 QY 118 CAAGCGTGAAGGATCAATGCTGTAGCAATGCTCATCAAGATACAGACCAAGGCGTATGT 177
 DB 180 TPAAGGAGGAGGATCAACGCCGAACAAATTGTTATCAAGATTAACGATCAAGGTTATGT 239
 QY 178 CACTTACATGCGCAACCACTATCATTTATTAAGTGTGTTCTTATGACGCTATCAT 237
 DB 240 GACCTCTATGAGAACCATTAATCATTAATGAGCAAGTTCTTATGATGACCATCAT 299
 QY 238 CAGTGAAGATTACTCATGAAGAAATCAAACTATAAGCTTAAAGATGAGATATTGTTAA 297
 DB 300 CAGTGAAGGCTCTTCATGAAGAAATCCGAATTATCAGTTGAAGATTCCAGACATTGTCA 359
 QY 298 TGAGGTCAAGGGTGTATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAGA 357
 DB 360 TGAATTCAGGGTGTATGTTATGTTATGATTAACGGTAAATCTATGTTTACCTTAAGA 419
 QY 358 TGCTGCCCAACGCGGATTAAGTCCGTATCAAAAAGAGAAATCAATCAACAAAACAAGAGA 417
 DB 420 TGCRCTCATGCGGATTAATATTCGAGCAAAAAGAGATTAAACGTCAACAAACAGAGAG 479
 QY 418 TAGTCAACATCGTGAAGGTGAATCTCAAGAAACGATGCTGTTGCTTGGACGTTTC 477
 DB 480 CAGTCATATATCAT-----AAGTCAGAGCAGATTAATGCTGTTGCTGACGACGAGAC 530

QY 478 GCAGAGCGCTATACAGATGATGATTATATCTTTAATGCTCTGATATCATAGAGA 537
 DB 531 CCAAGAGCTTATACAGAGATGATGGGTATATCTTCAATGATCATGATATCATTTAGAGA 590
 QY 538 TACTGTGATGCTTATATCTGTTCTCTGATGAGATCATTAACCTTACATTCTTAAGATGA 597
 DB 591 CACGGGTATGCTTATATCTGTTCTCTGATGAGATCATTAACCTTACATTCTTAAGATGA 650
 QY 598 GTTATCAGTACGAGGTTGGCTGTCGAGAAAGCTTCCATCTGCTGAGAAATCTGTC 657
 DB 651 GTTATCAGTACGAGGTTGGCTGTCGAGAAAGCTTATGGAATG----- 695
 QY 658 AATTCAGAACTTATCCGCCGCAAAAATAGCATTAACCTTCAAGAAACAACTGGGTACC 717
 DB 656 -GAAGCAGGATCTGCTCTCTTCAATTTAGTTAATAGCAATCCAGTCAACCA 754
 QY 718 TTCTGAAGCAATCCAGGATCTCAAAATTAACACAAAGCAACAAAGACACTTAACG 777
 DB 755 GATTGTCAGAGAACCAATCT-----GACTGTCACCTCAACTTA 794
 QY 778 TCAACGACGCTAAAGTATGATGATAGTCTCTTGAACAGCTCTCAAACTGCTTT 837
 DB 795 TCATCAAAATCAAGGGGAAAAACATTTCAAGCTTTTACGTGATGATGCTAAACCTT 854
 QY 838 GAGTCAAGACATGATGATCTGATGAGCTTGTCTTTGATCCAGACCAAAATCACAAGTGC 897
 DB 855 ATCAAGACCATGATGATCTGATGAGCTTATTTTCAACCGAGGCAAAATACAAAGTGC 914
 QY 898 AACAGCTAGAGGTGTGCAAGTGCACACGAGATCATTAACACTTATCTTACTCTTCA 957
 DB 915 AACCCGAGAGGTGATGCTGCTCCATGATGATCAATTAACACTTATCTTATGAGACA 974
 QY 958 AATGCTCAATTTGAGAGAACAAATGCTGCTGATTTTCCCTGCTTATCTGTTCAACA 1017
 DB 975 AATGCTCAATTTGAGAGAACAAATGCTGCTGATTTTCCCTGCTTATCTGTTCAACA 1034
 QY 1018 TTGGGTACAGATTCAAGGCGAAGCAACCAAGTCCACCAACGCTCCGGAACCTAGTCC 1077
 DB 1035 TTGGGTACAGATTCAAGGCGAAGCAACCAAGTCCACCAACGCTCCGGAACCTAGTCC 1094
 QY 1078 AAGCCCGAACCTGACCAAAATCTTAAATAGACTCAAAATCTTTGGTATGACGT 1137
 DB 1095 AAGTCCGCAACAGGCTCCAGCAATCAATGATGAGAA-----ATTGCTCAAGAAAGC 1148
 QY 1138 GGTACGAAAGTGGGGAAGATATGATTAATGCAAGAAAGGCACTCTCGTTATGCTTT 1197
 DB 1149 TGTGGAAGAGTGGGATGATGATGCTTTGAGAGAGATGAGTTTCTGTTATATCCC 1208
 QY 1198 TGCGAAAGATTTCACATCTGAAGCTGTTAAATCTTGAAGCAAGTATCAAAACAGA 1257
 DB 1209 AGCCAGAGATCTTTAGCAGAAACAGCAGCAGGCTTATAGCAAACTGGCCAGACAGA 1268
 QY 1258 GAGTGTTCACACACTTAACTGCTTAAAGAAAGAAATGTTGCTCTGTCGACCAAGATT 1317
 DB 1269 AAGTTATCTATAGTAGTAGAGAACTAAGAAACGACCTCCATCTAGATGAGGAATT 1328
 QY 1318 TTATATTAAGCATTAATCTGTTTACTGAGGCTATAAAGCTTGTGTTGAAATPAAGG 1377
 DB 1329 TTACATTAAGGCTTATGACTTACTAGCAAGATTCACCAAGTTTACTGATTAATAAGG 1388
 QY 1378 TCGTAATCTGATTTCAAGCCTTAGACAATTAATGAAGCTGATGATGATGATGATGATG 1437
 DB 1389 TCGACAAGTATTTTGGAGCTTTGATGATGATGATGATGATGATGATGATGATGATG 1448
 QY 1438 TAATTAAGAAATTTGATGATGATTTATGATGATGATGATGATGATGATGATGATGATG 1497
 DB 1449 TGATAAGTCAAGTAGTAGAGATATCTGCTTCTTGAAGTCTGATGATGATGATGATGATG 1508
 QY 1498 GCGACTTGGCAACCAATTTCTCAATTTGATGATGATGATGATGATGATGATGATGATG 1557
 DB 1509 ACGTTTGAAGAAACCAATTTGCGCAATTAATCTTACATGATGATGATGATGATGATG 1568
 QY 1558 ATTAGCTATAGTATACAGCTGATGATGATGATGATGATGATGATGATGATGATGATG 1617

```

Db      1569  GTTGGCAGGCAAGTACACAGCAGAGACGGTTATATCTTTGATCTCGTATATATACAG 1628
Oy      1618  TGATGAAGAGATGATATGTAACCCCTCATATGGCCATAGTCTCTGATTTGAAAAA 1677
Db      1629  TGAATGAGGGGATGCTATATGTAATCTCAATATGACCCATGCGCATGATTTAAAAA 1688
Oy      1678  TAGCCTTTCTGATAGAAAAAGTTGACGCTCAAGCCTATATCTAAAGAAAAAGTATCCT 1737
Db      1689  TAGTTTGTCTAGCTAGAGAGAGCGGCAAGCCAGCTTATGCTRAAGAAAAAGTTTAC 1748
Oy      1738  ACTTCATCTCCAGACGAGATGTTAAAGCAATCCAACTGAGATAGTGCAGACGTAT 1797
Db      1749  CCTCCTTCGACAGACCATCAGATTCAGAAATACTAGAGCAAAAAGAGCAGAACTAT 1808
Oy      1798  TTACATATGCTGAAAGGGAAAAACGAATTCATCTGCTTGCATCTCATATATGTTGA 1857
Db      1809  CTACAAACCGMGTGAAGAGAGCTTAAGAGAGTGCCTTATGCTATGCTTCAATCTTCA 1868
Oy      1858  GCATACAGTTGAGGTTAAAAACGGTATTTGATTTCTCTCATAGATCATTTACATTA 1917
Db      1869  ATATATCTAGATGATCAAAAAAGGATTTTATCTATCTCATTTATGACCATTTACATA 1928
Oy      1918  TATTAATTTTCTGTTGATGATCACAATACAAAGCTCCAAATGGCTATACCTTGA 1977
Db      1929  CATCAATTTGATGTTTGAAGAGGCTTTATGAGCACCTTAAGGGTATATCTTGA 1988
Oy      1978  AGATTTGTTTGGCAGATTAAGTACTAGTATGATACACCTGACAAAGCTCCATTTCTAA 2037
Db      1989  GGATCTTTTGGCGATGTCATGATCTATGATGAAATCAAAACGAACTCCGATTCGA 2048
Oy      2038  TGATGATGAGGCAATGCCAGTACATGTTAGTGAAGAAAGCACAAGTGAAGATCC 2097
Db      2049  TAAATGTTTGTGTAACGCTAGCGACCATGTTCAAGAAACAAAATGCTCAAGCTGATAC 2108
Oy      2098  AAATAGAACTTCAAGCGGATGA----- 2121
Db      2109  CAATCAACGGAAMAAACAGAGAGAGAAACCTCAGACAGAAAACTGAGGAAGAAAC 2168
Oy      2122  -----AGAGCCAGT 2130
Db      2169  CCTCGAGAGAGAAACCGCAAGGAGAAACAGAGTCTCCAAAAACACAGAGAGAAAC 2228
Oy      2131  AGAGAAACACCTGCTGAGCAGAAAGTCCCTCAATAGAGACTGAAAAAGTAAAGCCCA 2190
Db      2229  AGAAGATCACCAAGAGATCAGAAAGACCTCAGTCAAGTCAAAAGTTGAAGAAAA 2288
Oy      2191  ACTCAAGAACCAAGATTTTCTTGGCAAAAGTAAAGGATTTCTAGTCTGAAAGCCATGC 2250
Db      2289  ACTGAGAGAGGCTGAAGATTTACTTGGAAAAATCCAGATCCAAATTTATCAAGTCCATGC 2348
Oy      2251  AAGCAAACTCTAGTGTGTTTACGAAATTAATTTGACTCTTCAATATTAAGATTAATAG 2310
Db      2349  CAAAGAGCTCTCAGAGATTTAAAAAATTAATTTTATTTGACCCAGGACCAACAAATAC 2408
Oy      2311  TATCATGCGCAGAGCAGAAAAATTACTTGGTGTGTTAAAGGAAGTAA 2358
Db      2409  TATTATGCGAAGCTGAAAACTATTGCTTTTATTAAAGAGAGATTA 2456

```

RESULT 14
ABX06705
ID ABX06705 standard; DNA; 2517 BP.

AC ABX06705;

DT 11-FEB-2003 (first entry)

DE S. pneumoniae type 4 strain coding region #993.

KW Gene; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
ear infection; antiinflammatory; antibacterial; immunostimulant;
auditory; respiratory; gene therapy; vaccine.

```

XX      OS      Streptococcus pneumoniae type 4 strain.
OS      WO200277021-A2.
XX      PD      03-OCT-2002.
XX      PF      27-MAR-2002; 2002WO-1B02163.
XX      PA      27-MAR-2001; 2001GB-0007658.
XX      PA      (CHIR-) CHIRON SPA.
XX      PA      (GENO-) INST GENOMIC RES.
PI      Masignani V, Tettelin H, Fraser C;
XX      WPI; 2003-040579/03.
DR      P-PSDB; ABU01418.
XX      PT      New proteins and nucleic acid molecules from Streptococcus pneumoniae,
PT      useful as medicaments for treating or preventing a disease or infection
PT      due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
PT      or ear infection
XX      Claim 6; SEQ ID No 1985; 56bp; English.
XX      The invention relates to a protein comprising or having at least 50%
XX      identity to any of the 2469 amino acid sequences, identified in the
XX      specification (available on a computer readable format), or its fragment,
XX      expressed from 2469 of 2489 identified DNA coding regions from the
XX      Streptococcus pneumoniae type 4 strain genomic sequence appearing as
XX      AB556454. Also included are an antibody which binds one of the
XX      proteins, treating a patient by administering the protein, DNA or
XX      antibody (in a composition), a kit comprising first and second primers,
XX      which are the nucleic acid cited above or fragments between nucleotides
XX      8-100 of a sequence not defined in the specification, for amplifying a
XX      target sequence contained within a Streptococcus nucleic acid sequence,
XX      where the first primer is substantially complementary to the target
XX      sequence and the second primer is substantially complementary to the
XX      complement of the target sequence, and where the parts of the primers
XX      having substantial complementarity define the termini of the target
XX      sequence to be amplified, assay comprising contacting a test compound
XX      with the protein, and determining whether the test compound binds to the
XX      protein and a Streptococcus pneumoniae bacterium, where one or more
XX      genes encoding the proteins has been rendered inactive. The proteins,
XX      nucleic acid molecules, antibody and compositions are useful as
XX      medicaments for treating or preventing a disease or infection due to
XX      streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
XX      sepsis, otitis media or ear infection. They are also useful in developing
XX      vaccines, diagnostics and antibiotics. The methods are useful for
XX      identifying immunodominant proteins. The present sequence is one of
XX      the 2489 identified coding region from the genomic sequence.
XX      Note: The sequence data for this patent did not form part
XX      of the printed specification, but was obtained in electronic
XX      format directly from WIPO at
XX      ftp.wipo.int/pub/published_pct_sequences.
SO      Sequence 2517 BP; 830 A; 509 C; 545 G; 633 T; 0 other;

```

Query Match 41.4%; Score 990; DB 25; Length 2517;
Best Local Similarity 68.1%; Pred. No. 1.5e-239;
Matches 1478; Conservative 0; Mismatches 636; Indels 57; Gaps 5;

```

Oy      1  TTCTTACGAGTTGGAGCTGATATCAAGTAGAAGCGTTAAGAAAA---TAATGCTGTTTC 57
Db      60  TTCTTATGAACCTTGCTGCTACCAAGCTGCTGACAGTTAAGAAAGCTATATCAAGTTTC 119
Oy      58  CTATATAGATGAAGAAACAGAGCAGCAAAAACGAGAAATTGCTCCTGATGAGGTAG 117
Db      120  TTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 179
Oy      118  CAAAGCTGAAGATCAATGCTGAGCAAAATGCTCATCAAGATTAACAGACCAAGCTATGT 177

```

```

Db      180 TAAAGGAGGGGATCAACGCCGAACAAATCGTCATCAAGATTACGGAATCAAGTTATGT 239
Qy      178 CACTTACATGGGAGACCACTATCATATTATACAAATGGTAAAGGTTCTTATGACGCTATCAT 237
Db      240 GACCTCTCATGGAGACCACTTATCATTAATGACAAAGTCCCTTAATGATGCCATCAT 299
Qy      238 CAGTGAAGAATTACTCATGAAGAATCCAAACTTAAGCTAAAAGATGAGGATATTTGTTAA 297
Db      300 CAGTGAAGAGCTCCCTCATGAAGAATCCGAATTATCAGTTGAAGGATTCAGACATTTGCA 359
Qy      298 TGAAGTCAGAGGTGATGATTATCAAGAGTAAATGAAAAATACTATGTTTACCTTAAGA 357
Db      360 TGAATTCAGAGGGGTGTTATGTTATCAAGGTAGATGAAAAATACTATGTTTACCTTAAGA 419
Qy      358 TCGTCGCCAGCGGATPAAGTCGCTCAAAAAGGAAATCAATCCGCAAAAACAAGAGA 417
Db      420 TCGAGCTCATGGGATPAATATTGGAACAAAAGAGATTAAACGTGACAGAGAGAGA 479
Qy      418 TAGTCAACATCGTGAAGTGAACCTCCAAAGAAACGATGGTGTGCTTGCTTGCACTTC 477
Db      480 CAGTCAATATCACGGGGGTGGTCT-----AACGATCAAGCAGTGTGACGCCAGAGC 533
Qy      478 GCAAGAGCCCTATATCAAGATGATGATTATCTTAAATGCTTCTGATATCATAGAGA 537
Db      534 CCAAGGACCTATACCAACGAGATGATGTTATCTTCAATGCAATCGATATCATAGAGA 593
Qy      538 TACTGGTATGCTTATATGTTCTCATGAGATCATTAACATTAATCTTAAAGATA 597
Db      594 CACGGGTATGCTTATATGTTCTTACGCGACCAATACATTAATCTTAAAGATA 653
Qy      598 GTTATCAGCTAGCAGATGGCTGCTGACAGAAAGCCTTCTATCTGATCGAGAAATCTGTC 657
Db      654 GTTATCAGCTAGCAGATGGCTGCTGACAGAAAGCCT----- 688
Qy      658 AAATTCAGAACCTATCGCCGACAAATATGCGATTAACCTTCAAGAACTGGGTACC 717
Db      689 -----ATTGGAATGGAGAGAGGATCTCGTCTTCTTCAAGTTCTAGTTAATATGC 740
Qy      718 TTCTGTAACCAATCCAGAACTACAAATCTAACACAGCAACAAGACAACCTAACG 777
Db      741 AAATCAGCTCAACCAAGATTGTACAGAAACCAATCTGACTGTACTCCAACTTATGA 800
Qy      778 TCAGCAAGTCAAGATATGACATTGATGCTCTTGAACACAGCTTACAACTGGCTTT 837
Db      801 TCA-----AAATCAAGGGGAAAACTTTCAGCCTTTTACGTAATGTAATGCTTAACTCT 857
Qy      838 GAGTCAAGCAGATGTAAGATCTGATGAGCTTGTCTTGTATCCAGACAATATCAAGATGC 897
Db      858 ATCAGAAAGCCATGTGATCTGATGAGCTTATTTTCGACCCAGCGCAATATCAAGATGC 917
Qy      898 AACAGCTAAGGTGTGCAAGTGCACACGAGATCATTAACACTTATCCCTTACTCTCA 957
Db      918 AACCGCAGAGGTGATGCTGCTCCCTCATGTAAACATTAACACTTATCCCTTATGAMA 977
Qy      958 AATGCTGATTTGAAGAAAGAAATCGCTGTAATTTTCCCTTCCCTTCTGTTACAAACA 1017
Db      978 AATGCTGATTTGAAGAAAGAAATCGCTGTAATTTTCCCTTCCCTTCTGTTACAAACA 1037
Qy      1018 TTGGGTACAGATTCAAGGCCAGAACCAACCAAGTCCAAACGACTCCGGAACCTAGTCC 1077
Db      1038 TTGGGTACAGATTCAAGGCCAGAACCAACCAAGTCCAAACGACTCCGGAACCTAGTCC 1097
Qy      1078 AAGCCCGCAACTGCACCAATCTTAAATAGACTCAATTTCTC-----TTT 1125
Db      1098 AAGTCGCAACCTGCACCAATCTTAAACCAAGCTCCAAACCAATTCAGATGAGAAAT 1157
Qy      1126 GGTAGTCACTGTGTCAAAAAAGTTGGGAAAGATATGATTGCAAGAAAAAGGACATCTC 1185
Db      1158 GGTCAAAAGAGCTGTTCAAAAAGTAGGCAATGTTATGCTTTGAGAGAAATGAGATTTC 1217
Qy      1186 TCGTTATGCTTTGCGAAAAGATTACCATCTGAAACTGTTAAATCTTGAAAGCAAGTT 1245
Db      1218 TCGTTATATTCAGCCAGAGATCTTTCAGACAGAAACAGACAGCGCATTTGATGCAAACT 1277

```

```

Qy      1246 ATCAAAACAGAGAGTGTTCACACACTTAACTGCTTAAAAAAGAAATGTTGCTCTCG 1305
Db      1278 GGCACAGCAGAAAGTTATCTCATATAGCTAGAGACTAAGAAACTGACCTCCCATCTG 1337
Qy      1306 TGAACAGAAATTTATGATTAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGT 1365
Db      1338 TGATCGAGAAATTTACAAATTAAGCTTATAGACTTACTAGCAAGAAATTCACCAAGATTACT 1397
Qy      1366 TGAATAAAGGTGCTAATTCGATTTCCAGCCTTAGCAAAATTTATTAAGACGCTTGA 1425
Db      1398 TGATATTAAGGTGCAACAGTTGATTTGAGGCTTTGATTAACCTGTTGAAACGACTCA 1457
Qy      1426 TGATATGATGATTAATAAAGAAAAATGAGTAGATGTTATGGAATCTTACGACCAAT 1485
Db      1458 GGAATGCCAAGTAAATTAAGTCAAGTTGATGATGATTTCTTGCTCTTATGCTCGAT 1517
Qy      1486 TACCATCCAGAGCACTTGGCAACCAATCTCAAAATGAGTATGATGAAGAGCAAGT 1545
Db      1518 TCGTATCCAGAAACGTTTAGAAAAACCAATGCGCAATTTACCTACACTGATGAGAT 1577
Qy      1546 TCGTATGCTCAATTAGCTGATTAAGTATTAACAGTCAAGTGTACATTTTGAATGA 1605
Db      1578 TCAAGTAGCAAGTTGGCAGGCAAGTACACAACAGAAACGGTTATATCTTGTATCTCG 1637
Qy      1606 TGATATTAATCAGTATGAAGAGATGCAATGTAACGCTCATATGGGCAATGCACTG 1665
Db      1638 TGATATTAACAGTATGAGGGGATGCTTATTAATCACTCAATTAAGACCAATGCCACTG 1697
Qy      1666 GATTGAAAAAGATAGCCTTTCTGATTAAGAAAAAGTTGAGCTCAAGCCTATATTAAGA 1725
Db      1698 GATTAAAAAAGATAGTGTGTGTAAGCTGAGAGAGGGGAGCCAGGGCTTATGCTAAGA 1757
Qy      1726 AAAAGTATCTTACTCTCATCTCCAGACGCAAGTGTAAAGCAATCCAACTGAGATG 1785
Db      1758 GAAAGTGTGACCCCTCTTCCAGACACCATCAGATTCAGAAAAATACGAGCAAAAG 1817
Qy      1786 TGCAGAGCTATTTCAATCGTGTAAAGGGGAAAAACGAATCCACTGCTGCACTTCC 1845
Db      1818 AGCAGAGCTATCTCAACCGGTAAAGACACTTAAGAAAGTGCCACTGATGATGTGCC 1877
Qy      1846 ATATATGTTGAGCATACAGTTGAGGTTAAAAACGGTAAATTTGATTATCTCATTAAGA 1905
Db      1878 TTACAAATCTCAATATATCTGTAAGATGAAAAACGGTATTTAATCATCTCATTTATGA 1937
Qy      1906 TCATTACCAATATTAATTTGCTTGTGTTGATATCAACATACAAAAGCTCCAAATGG 1965
Db      1938 CCATTACCAATTAATTAATTTGAGTGTGTTGACAAAGGCTTTATGAGGCACTTAAGGG 1997
Qy      1966 CTATACCTTGAAGATTTGTTTGGACGATTAAGTAACTAAGTAAGAACCCCTGAGGAAG 2025
Db      1998 GTATACCTTGAAGATCTTTTGGCGACGTCAAGTAACTAATGTGCAACATCCAAAGAG 2057
Qy      2026 TCCACATTTCAATGATGATGAGGGAATGCCAGTAGACATGTGTTAAGCAAGAAAGACA 2085
Db      2058 TCCGATTCAGATTAATGTTTGGTATGACGCTTACGACATGTTCCGTAATAAAGTATGA 2117
Qy      2086 CAGTGAAGTCCAAATTAAGAACTTCAAAAGCGATGAAGCAAGTAGAGAAAAACCTGTC 2145
Db      2118 CCAAGACAGTAAACCTGATGAAGATPAAGGAACATGATGAAGTATGAGCCAACTCAC 2177
Qy      2146 TGAAGCAGAAAG 2156
Db      2178 TGAATCTGATG 2188

```

RESULT 15
 AAA47602
 ID AAA47602 standard; DNA; 2531 BP.
 XX
 AC
 XX
 DT 20-OCT-2000 (first entry)

XX XX Recombinant variant of Sp36 gene (Sp36D) of *S. pneumoniae*.
 DE Streptococcus pneumoniae; infection; vaccine; coiled coil region;
 XX histidine triad residue; Sp36; antibody; otitis media;
 KM nasopharyngeal infection; bronchitis; sepsis;
 KW meningitis; lobar pneumonia; ds.
 XX OS Streptococcus pneumoniae.
 XX FH Key Location/Qualifiers
 FT CDS 1..2517
 FT /*tag= a
 FT /product= Sp36D polypeptide
 XX WO200037105-A2.
 XX PD 29-JUN-2000.
 XX PF 21-DEC-1999; 99WO-US30390.
 XX PR 21-DEC-1998; 98US-0113048.
 XX PA (MEDI-) MEDIMUNE INC.
 XX PI Johnson LS, Koenig S, Adamou JE;
 XX DR WPI: 2000-452129/39.
 XX DR P-PSDB; AAB01466.
 XX PT Vaccine useful for prophylaxis and treatment of pneumococcal infections
 PT such as otitis media, nasopharyngeal and bronchial infections,
 PT comprises Streptococcus pneumoniae proteins
 XX PT comprises Streptococcus pneumoniae proteins
 XX PS Disclosure; Page 57-58; 70pp; English.
 XX CC Although a number of proteins have been suggested as being involved
 CC in the pathogenicity of Streptococcus pneumoniae, there still remains
 CC a need to identify polypeptides having epitopes in common from
 CC various strains of *S. pneumoniae* in order to utilize such
 CC polypeptides in vaccines to protect against a wide variety of
 CC *S. pneumoniae*. New vaccine compositions are described which comprise a
 CC streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino
 CC acids in length that comprise at least one histidine triad residue
 CC (HxxHxx) or a coiled-coil region, or an antibody directed against
 CC these features. The vaccine is useful in protecting against infection
 CC by Streptococcus pneumoniae. The vaccine composition comprising
 CC antibodies to its useful for passive immunization for treating
 CC pneumococcal infections which includes otitis media, nasopharyngeal
 CC and bronchial infections.
 CC CC
 CC CC
 SO Sequence 2531 BP, 836 A, 511 C, 547 G, 637 T; 0 other;
 Query Match 41.4%; Score 990; DB 21; Length 2531;
 Best Local Similarity 68.1%; Pred. No. 1.5e-239;
 Matches 1478; Conservative 0; Mismatches 636; Indels 57; Gaps 5;

QY 238 CAGTGAAGAAATTAATCAATGAAAGATCCAACTATAGCTAAAGATGAGATATTGTTAA 297
 DB |||||
 QY 300 CAGTGAAGAGCTCCTCATGAAGAAGATCCGAATTAATCAGTTGAAGATTGATGCA 359
 DB |||||
 QY 298 TGAAGTCAAGGGTGAATATGTTATCAAGATGATGAGAAATATCATATGTTTACCTTAAGCA 357
 DB |||||
 QY 360 TGAATCAAGGGTGAATATGTTATCAAGTGAAGTGAAGAAATATCATATGTTTACCTTAAGCA 419
 DB |||||
 QY 358 TGCTGCCAGGEGATACGTCCTACAAAGAGAAATCAATGACAAAGAAACAAAGACA 417
 DB |||||
 QY 420 TGCAGCTCATGCGATTAATATTCGACAAAGAAAGATTAAGCTCGAAGACGAGACA 479
 DB |||||
 QY 418 TAGTCAACATCGTGAAGGTGAAGTCCCAAGAAAGATGCTGTGCTGGACGTTTC 477
 DB |||||
 QY 480 CAGTCAATATCAGCGGGGTGTTCT-----AAGCATCAACGATAGTTGACGACGAGGC 533
 DB |||||
 QY 478 GCAGAGACGTAATCACTACAGATGATGTTATATCTTTAATGCTTGTGATATATAGAGA 537
 DB |||||
 QY 534 CCAAGGACGCTATCAACGAGATGATGTTATATCTTCATGATCATCTGATATCATTTGAGAGA 593
 DB |||||
 QY 538 TACTGCGATGCTTAATATCGTTCTCCTCATGAGATCATTAACATTAATCTCTAAGAAATGA 597
 DB |||||
 QY 594 CACGGTGAATGCTTAATATCGTTCTCAGCGACCATTAACATTAATCTCTAAGAAATGA 653
 DB |||||
 QY 598 GTTATCAGCTAGCGAGTTGCTGCTGAGAAAGCTTCTATCTGTGAGGAGAAATCTGTTC 657
 DB |||||
 QY 654 GTTATCAGCTAGCGAGTTGCTGCTGAGAAAGCTTCTATCTGTGAGGAGAAATCTGTTC 688
 DB |||||
 QY 658 AAATTCAGAAACCTATGCGCCGACAAATAGGATPAACCTTCAAGAACAACTGGGTAAC 717
 DB |||||
 QY 689 -----ATTGGAATGGGAAGACAGGATCTGCTCTTCTTCAAGTTCTATGTTAATGC 740
 DB |||||
 QY 718 TTCTGTAGCAATCAGAGAACTACAAATATCTAACAAGCAACAACGACACCTAAGACAG 777
 DB |||||
 QY 741 AAATTCAGCTCAACCAAGATTTGTCAGAGAACCAATCTGCTCATCTCAACTCACTTACA 800
 DB |||||
 QY 778 TCAAGCAAGTCAAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 837
 DB |||||
 QY 801 TCA---AAATCAAGGGAAGAAATTTTCAAGCTTTTACGTAATGTTATGTTAAACCTT 857
 DB |||||
 QY 838 GAGTCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 897
 DB |||||
 QY 858 ATCAAGAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 917
 DB |||||
 QY 898 AACGCTAGAGTGTGAGAGCCACACGAGATCATTAACATCACTTCACTTCACTTCACTTCA 957
 DB |||||
 QY 918 AACGCTAGAGTGTGAGAGCCACACGAGATCATTAACATCACTTCACTTCACTTCACTTCA 977
 DB |||||
 QY 958 AATGCTGAATTTGGAAGAACCAATGCTGATTAATTTCCCTTGTATCGTTCAACCA 1017
 DB |||||
 QY 978 AATGCTGAATTTGGAAGAACCAATGCTGATTAATTTCCCTTGTATCGTTCAACCA 1037
 DB |||||
 QY 1018 TTGGGTACAGATTTCAAGGCGCAGAACCAACAGTCCACACGACTCCGAACTGATGC 1077
 DB |||||
 QY 1038 TTGGGTACAGATTTCAAGGCGCAGAACCAACAGTCCACACGACTCCGAACTGATGC 1097
 DB |||||
 QY 1078 AGGCGCCGACACGCGACCAATCTTAAATAGACTCAAAATTCCTC-----TTT 1125
 DB |||||
 QY 1098 AAGTCCGAACTTCCACCAATCTTCAACGCTTCAACGATCAATCAATGATGAGAAAT 1157
 DB |||||
 QY 1126 GGTATGATGCTGTATGAGAAAGTTGGGGAAGATATGATTTGCAAGAAAGAGGATCTC 1185
 DB |||||
 QY 1158 GGTCAAGAAAGCTGTTGGAAGAAAGTGAAGCGATGATGATGATGATGATGATGATGATGATGAT 1217
 DB |||||
 QY 1186 TCGTTATGCTTTTTCGAAAGATTTTACATCTGAAATCTGTTAAATCTTGAAGCAAGTT 1245
 DB |||||
 QY 1218 TCGTTATGCTTTTTCGAAAGATTTTACATCTGAAATCTGTTAAATCTTGAAGCAAGTT 1277
 DB |||||
 QY 1246 ATCAAAACAGAGAGTGTTCACACATTTTACATCTGAAATCTGTTAAATCTTGAAGCAAGTT 1305
 DB |||||
 QY 1278 GGCACACAGAGAGAGTGTTCACACATTTTACATCTGAAATCTGTTAAATCTTGAAGCAAGTT 1337
 DB |||||
 QY 1306 TGACCAAGAAATTTATGATAAAGCATTAATCTGTTAACTGAGGCTCATAAAGCTTGT 1365

DB 1338 TGATGAGAAATTTTCAATTAAGGCTTATGACTTACTAGAGAGAAATTCACCAAGATTACT 1397
 OY 1366 TGNAAATATAGGCGGTAATTTGATTCGAGCGCTTAGCAATTAATTAAGAGCGCTGAA 1425
 DB 1398 TATATATTAAGTTCGACCAAGTTGATTTGAGCTTGGATTAACCTGTTGAAAGACTCA 1457
 OY 1426 TGATGATGCTAATTAATTAAGAAATTTGATGATGATTTATTTGCACTTCTTGAACCAAT 1485
 DB 1458 GGATGTCGCAAGTAAAGTCAAGTTAGTATGATGATTTCTTGGCTTCTTGGCTGCAAT 1517
 OY 1486 TACCCATCCAGAGCGACTTGGCAACCAATTTCTCAATTTAGTATGATGAGAGCAAGT 1545
 DB 1518 TCGTATCCAGAACGTTTGAAGAAACCAATGCGCAAAATTAACCTCACTGATGATGAGAT 1577
 OY 1546 TCGTATTCCTCAATTTGCTGATTAAGTATCAACGTCAGATGGTTATATTTTGTAGTAA 1605
 DB 1578 TCAAGTAGCAAGTTGGCAGGCAAGTACCAAGAAAGCGTTATATCTTGTATCTCTG 1637
 OY 1606 TGATATTAATCAAGTATGAGAGATGATGATGATGATGATGATGATGATGATGATGATG 1665
 DB 1638 TGATATTAACCAAGTATGAGAGGAGATGCTATGATGATGATGATGATGATGATGATGATG 1697
 OY 1666 GATTGGAAGAAATAGCTTTCTGATTAAGAGAAAGTTGCACTCAAGCTTATATTAAGA 1725
 DB 1698 GATTAAAAAGATAGTTGCTGAAAGCTGAGAGAGCGGAGCGGCTTATGCTTAAGA 1757
 OY 1726 AAAAGGATCTTACCTCCATCTCCAGACCGCAATGTTAAAGCAATCCAACTGAGATG 1785
 DB 1758 GAAAGGTTGAGCCCTCTTCAACAGACATGAGATGAGAAATATGAGGCAAAAG 1817
 OY 1786 TGCACAGCTATTTTCAATCTGCTGTAAGAGGAGGAAACCAATTTCCACTGCTTCACTCC 1845
 DB 1818 AGCAGAAGCTATCTACACCGGCTGAAGACGCTAAGAGGTCCTGATGATGATGATGATG 1877
 OY 1846 ATATATGCTGAGCATCAGTTGAGTTAAACCGTAAATTTGATTTCTTCATTAAGA 1905
 DB 1878 TTACATCTCTCAATATCTAGTAAAGTCAAAACCGTAAATTTATCATCTCATATTA 1937
 OY 1906 TCATTAACCAATATTAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATG 1965
 DB 1938 CCATTAACCAATATCAATTTGAGTGAAGCTTAAAGAGGCTTAAAGAGGCTTAAGAGG 1997
 OY 1966 CTATACCTTGAAGATTTGTTGCGAGCATTAAGTACTACTAGTAAACACCTGACGAAAG 2025
 DB 1998 GATATCTCTGAGATCTTTTGGCAGCTCTCAAGTACTATGTCGAACATCCAAAGCAAG 2057
 OY 2026 TCCACATTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2085
 DB 2058 TCCGATTCAGATTAATGTTTGGTAAAGCTGACGATGTTGCTGAAAAATTAAGTAA 2117
 OY 2086 CAGTGAAGATCCAAATTAAGAACTTCAAGCGGATGAAGCGCAAGTAAAGAAACACCTGC 2145
 DB 2118 CCAAAACAGTAAACCTGATGAAGATTAAGAACATGATGAAGTGAAGCAACTCACCC 2177
 OY 2146 TGAGCCAGAG 2156
 DB 2178 TGAATCTGATG 2188

RESULT 16
 ABS56454
 ID ABS56454 standard; DNA; 2162598 BP.

XX 10-FEB-2003 (first entry)
 DT Streptococcus pneumoniae type 4 strain complete genome.
 XX ds; bacterial meningitis; pneumonia; sepsis; otitis media; genome;
 KM ear infection; antiinflammatory; antibacterial; immunostimulant;
 KW auditory; respiratory; gene therapy; vaccine.

XX OS Streptococcus pneumoniae type 4 strain.
 XX PN MO200277021-A2.
 XX PD 03-OCT-2002.
 XX PF 27-MAR-2002; 2002MO-IB02163.
 XX PR 27-MAR-2001; 2001GB-0007658.
 XX PA (CHIR-) CHIRON SPA.
 XX PA (GENO-) INST GENOMIC RES.
 XX PI Masignani V, Tettelin H, Fraser C;
 XX DR WPI; 2003-040579/03.
 XX PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 XX PT useful as medicaments for treating or preventing a disease or infection
 XX PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
 XX PT or ear infection -
 XX PS Claim 17; SEQ ID No 4979; 56bp; English.
 CC The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC ABS56454. Also included are an antibody which binds one of the
 CC proteins, treating a patient by administering the protein. DNA or
 CC antibody (in a composition), a kit comprising first and second primers,
 CC which are the nucleic acid cited above or fragments between nucleotides
 CC 8-100 of a sequence not defined in the specification, for amplifying a
 CC target sequence contained within a Streptococcus nucleic acid sequence,
 CC where the first primer is substantially complementary to the target
 CC sequence and the second primer is substantially complementary to the
 CC complement of the target sequence, and where the parts of the primers
 CC having substantial complementarity define the termini of the target
 CC sequence to be amplified, assay comprising contacting a test compound
 CC with the protein, and determining whether the test compound binds to the
 CC protein and a Streptococcus pneumoniae bacterium, where one or more
 CC genes encoding the proteins has been rendered inactive. The proteins,
 CC nucleic acid molecules, antibody and compositions are useful as
 CC medicaments for treating or preventing a disease or infection due to
 CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
 CC sepsis, otitis media or ear infection. They are also useful in developing
 CC vaccines, diagnostics and antibiotics. The methods are useful for
 CC identifying immunodominant proteins. The present sequence is the
 CC Streptococcus pneumoniae type 4 strain genome sequence.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 2162598 BP; 654373 A; 427176 C; 431369 G; 649680 T; 0 other;
 Query Match 41.4%; Score 990; DB 25; Length 2162598;
 Best Local Similarity 68.1%; Pred. No. 2.8e-238;
 Matches 1478; Conservative 0; Mismatches 636; Indels 57; Gaps 5;
 OY 1 TTCTTACGAGTTGGACCTGTATCAAGCTAGAACGTTTAAAGAAA--TATCGCTTTC 57
 DB 838008 TTCCATTAAGTATGCTGCTCAACCAAGCTGCTAGTTAAGAAAGTCAATCGAGTTTC 838067
 OY 58 CTATATAGTGAAGAAACCAAGCAAGCAAAACGAGATTTGACTCCGATGAGGTTAG 117
 DB 838068 TTATATATAGTGTATGATGCTGCTCAAGCAAGCAAAACCTTGACACCAAGATGATGAG 838127
 OY 118 CAAGCTGAAGAAATCAATGCTGAGCAAAATGCTCATCAAGATTAACGACCAAGGCTATGT 177
 DB 838128 TAAAGAGGAGGAGGATCAACGCGCAACCAATGCTCATCAAGATTAAGGATCAAGGTTATGT 838187

178 CACTACATGGCGCAGCTATCATTATTAACAAGTAAAGGTTCTTATGACGCTATCAT 237
838188 GACCTCTATGAGACCAATTATCATTATATG6CAAGGTCCTTATGATGCCATCAT 838247
238 CAGTGAAGAATTACTCATGAAGAAGATCCAAACTATAGCTAAAGAAGATGAGATATTGTTAA 297
838248 CAGTGAAGAAGCTCTCATGAAAGATCCGAATTATCAGTTGAAGATTCAGACATTTGTCA 838307
298 TGAGGTCAAGGTTGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACTTTAAGGA 357
838308 TGAATTCAGAGGTTGTTATGTTATCAGGTAGATGGAATAATCTATGTTTACTTTAAGGA 838367
358 TGTCGCCACGCGAGTAAAGTCCGACAAAGAGGAATCAATCGACAAAACAAGACA 417
838368 TGCACTCATGCGGATTAATTTCGACAAAGAGAGATTAAGAGCTCAGAAAGCAGAGACA 838427
418 TAGTCACATCGTGAAGGTGGAACCTCAAGAAACGATGCTGTTGCTTGGCAGCTTC 477
838428 CAGTCATATCACGGGGTGGTTC-----AACGATCAAGCAGTAGTTGCAGCCAGAGC 838481
478 GCAAGAGGCTTACTACAGATGATGTTTATCTTTAATGCTTGTGATATCATAGAGA 537
838482 CCAAGGAGCTATACAGAGATGATGTTATATCTTCAATGATCTGATATCATTAGAGA 838541
538 TACTGCTGATGCTTATCGTTCTCATGAGAGATCATTACCATTCATTCCTAAGAATGA 597
838542 CACGGGTATGCTTATCGTTCTCATGAGGACATTAACCTTATCTTCAAGAAATGA 838601
598 GTTATCAGCTAGCAGATTGGCTGCTGCAAGACCTTCTATCTGTGTCAGGAATCTGTC 657
838602 GTTATCAGCTAGCAGATTGCTGCTGCAAGACCT----- 838636
658 AAATTCAGAACCTATGCCCGACAAATAGCATTAACCTTCAAGAACAACTGGGTAC 717
838637 -----ATTGGAATGGAAAGCAGGATCTGCTCTTCTTCAAGTTTATGTTATATGC 838688
718 TTCTGTAAGCAATCAGGAACCTAACAATACTAACAGACAAACAGACAACTAAG 777
838689 AAATTCAGCTAACCAAGATTTGTCAGAGAACCAACATGACTGTCATCCAACTTATCA 838748
778 TCAAGCAAGTCAAAAGTATGAATGATGATGCTCTTGAACAGCTCTAACAACCTGCTT 837
838749 TCA---AAATCAGGGGAAACATTTCAAGCTTTTACGTGAATGTTGCTAAACCTT 838805
838 GAGTCACAGCATGATAGAACTGATGCTTGTCTTTATCCAGACAAATCACAAGTCG 897
838806 ATCAGAACGCGATGGAATCTGATGCTTATTTTCAGCCAGGCAAAATCACAAGTCG 838865
898 AACAGCTAGAGGTGTCAGTGCACACAGAGATCATTACCACTTACCCCTTACTCTCA 957
838866 AACGCCAGAGGTGATGCTGCTCCCTCATGTRAACATTAACCTTATCCCTTATGAAACA 838925
958 AATGCTGAATTTGAAGAAGCAATGCTGTAATTTATCCCTTCTGTTACCAACA 1017
838926 AATGCTGAATTTGAAGAAGCAATGCTGTAATTTATCCCTTCTGTTACCAACA 838985
1018 TTGGGTACCAATTTAAGGCCAGAAACAACCAAGTCCACACCGACTCCGGAACCTAGTCC 1077
838986 TTGGGTACCAATTTAAGGCCAGAAACAACCAAGTCCACCAATCGACTCCGGAACCTAGTCC 839045
1078 AGGCCCGCAACCTGACCAAAATCTTAAATAGACTCAATTCCTC-----TTT 1125
839046 AAGTCCGCAACCTGACCAAAATCTTCAACAGCTCAAGCAATCCAAATTTGATGAGAAAT 839105
1126 GGTATGACGTGTGACGAAAGATTGGGAAAGATATGTTATCGAAGAAAGGCGATCTC 1185
839106 GGTCAAAAGAGCTGTTGAAAGTAGGCGATGTTATGCTTTGAGGAATGAGGTTTC 839165
1186 TCGTTATGCTTTTGGGAAAGATTTACATCTGGAACCTGTTAAATCTTGAAGCAAGTT 1245
839166 TCGTTATGCTTTTGGGAAAGATTTTTCAGCAGAAACACAGAGGCAATTTGATGAGCAACT 839225

1246 ATCAAAACAGAGAGTGTTCACACACTTTAATCTGTAAAGAAAGAAATGTGCTCTCG 1305
839226 GCCAAGCAGAGAAAGTTATCTCTAATAAGCTAGAGCTAGAAAGAACTGACCTCCATCTAG 839285
1306 TGACCAAGAAATTTATGATTAAGCATATATCTGTTACTGAGGCTCATTAAGCCTTGT 1365
839286 TGATCGAAATTTTACATTAAGGCTTATGACTTACTGCAAGAAATTCACCAAGATTTACT 839345
1366 TGNAAATPAAGGTTGTAATCTGATTTCCAAAGCCTTGAACAAATTTATAGAACGCTTGA 1425
839346 TGATTAATPAAGTTGACAGATGATGTTTGAAGCTTTGATGATACCTGTTGAAACGACTCA 839405
1426 TGATGAATCGACTAATPAAGAAATTTGATGATATTTATGCAATTCCTAGCACAAT 1485
839406 GGATGTCCTCAAGTATTAAGTCAGATGATGATGATGATGATGATGATGATGATGATGAT 839465
1486 TACCATTCAGAGGACTTGGCAACCAATTTCCAAATTTGATGATGATGATGATGATGAT 1545
839466 TCGTCATCCAGAACGTTTGAAGAAACCAATTCGCAAAATTAACCTACACTGATGATGAT 839525
1546 TCGTATTTGCTCAATTAAGCTGATTAAGTATACAGCTGATGATGATGATGATGATGAT 1605
839526 TCAAGTACCAAGTTGCGAGCAAGTACACACAGAAAGGTTATATCTTTGATCTCG 839585
1606 TGATATATCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1665
839586 TGATATTAACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 839645
1666 GATTGGAAGATATGCTTCTTGTATGAAGAAAGTTGCACTGATGATGATGATGATGATGAT 1725
839646 GATTGGAAGATATGCTTCTTGTATGAAGAAAGTTGCACTGATGATGATGATGATGATGAT 839705
1726 AAAAGTATCCTTACCTCATCTCCAGACGAGATGTTAAAGCAATCCAACTGAGATAG 1785
839706 GAAGGTTTGAACCTCTCTTTCAGACAGCATCAGAGATTCAGAAATCTGAGGCAAAAG 839765
1786 TGACAGCAGTATTTAACAATGCTGTAAGGAGGAAAGCAATTCACACTGTTGCACTTC 1845
839766 AGCAAGAGCTATCTACAAACCGGTGAAGAGCTTAAGAGGTCCTGATGCTATGCTC 839825
1846 ATATATGCTGAGCATACAGTTGAGTTTAAAGGTTAAATTTGATTTTCTCTATAGGA 1905
839826 TTACAAATCTTCAAAATATCTGTAAGTCAAAAGGTAAGTTATCTTATCTCTATGTA 839885
1906 TCATTCATTAATTTTAAATTTGCTGTTGATGATCACAATTAAGGCTCCAAATG 1965
839886 CCATTCATTAATTTTAAATTTGCTGTTGATGATCACAATTAAGGCTCCAAATG 839945
1966 CTATACCTTGAAGATTTGTTGCGACATTAAGTATGATGATGATGATGATGATGATGATGAT 2025
839946 GTATACCTTGAAGATTTTGGCGACGTGCAAGTATGATGATGATGATGATGATGATGATGAT 840005
2026 TCACATTTCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2085
84006 TCCCATTTCAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840065
2086 CAGTGAAGATCCAAATPAAGAACTTCAAGCGGATGAAGAGCAGTGAAGAAACACCTGC 2145
84006 CCAAGACAGTAACTTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAA 840125
2146 TGAAGCAGAAAG 2156
840126 TGAATCTGATG 840136

RESULT 17
AAV27356
ID AAV27356 standard; DNA; 2290 BP.
XX AAV27356;
AC
XX
XX
XX
XX
02-Oct-1998 (first entry)
XX


```

QY 1306 TGACCAAGATTTTATGATAAAGCATATATCTGTAACTGAGGCTCATAAAGCCTTGT 1365
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1279 TGATCGAATTTTACATTAAGGCTTATGACTTACTACCAAGAAATTCACCAAGATTACT 1338
QY 1366 TGNMAATTAAGGTCGTAAATTCTGATTTCCAAAGCCTTACCAAAATTTATGAACGTTGAA 1425
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1339 TGATTAATTAAGGTCGAAAGTTGATTTGAGGCTTTGATTAACCTGTGGAACGACTCA 1398
QY 1426 TGATGAATCGACTAATAAGAAAAATTTGATGATTTATTTGACCTTCTAGACCAAT 1465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1399 GGAATGTCAGATGATTAAGTCAAGTCTAGTGAAGATATTTCTGCTCTTTAGCTCCGAT 1458
QY 1486 TACCATTCAGAGCAGCTTGGCAACCAATTTCTCAATTTAGTATCTGAAGAAGT 1545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1459 TCGTATTCAGAGCGTTTAGGAAAACCAATGCCAAATTTACCTACATCTGATGATGAT 1518
QY 1546 TCGATTTCTCAATTAGCTGATAGTATACACGTCAGATGTTACATTTTGTATGAACA 1605
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1519 TCAAGTAGCCAAAGTTGGCAGGCAAGTACACAAACAGAACGTTATATCTTTGATCCTCG 1578
QY 1606 TGATATATTCAGATGTAAGAGATGATGATGATGATGATGATGATGATGATGATGATG 1665
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1579 TGATATTAACAGATGATGAGGGAGTCTGATGATCTCACATATGACCCATAGCCACTG 1638
QY 1666 GATTGAAAAAGATAGCCTTTCTGATTAAGAAAAAGTTGACAGCTCAAGCCTTACTAAAGA 1725
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1639 GATTAAAAAAGATAGTTGTCTGAGCGTGAAGAGCGGACCCAGCGCTTATGCTAAAGA 1698
QY 1726 AAAAGTATCTTACTCCTCATCTCCAGAGCAGATGTTAAGCAATCCAACTGAGATAG 1785
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1699 GAAAGTTTGACCCCTCTCGACAGACCATGAGATTCAGAAATATCTGAGGCAAAAGG 1758
QY 1786 TGCGAGAGCTTTTATCATTCGTTGAAGGGGAAAAAGCAATTCACCTGTTCCACTCC 1845
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1759 AGCAGAGCTTATCTAACAACCGGTGAAGAGCTTAAGAGGCTGATGATGATGCC 1818
QY 1846 ATATATGTTGAGCATGATGAGTGAAGTAAAAACGGTATTTGATTTCTTCATTAAGA 1905
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1819 TTACAATCTTCAATATATCTGTAAGATCAAAAACGGTATTTAATCACTCATTAAGA 1878
QY 1906 TCATTTACATTAATTAATTTGCTGGTTGATGATCAACATCAAAAGCTCCAAATGG 1965
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1879 CCATTTACCATTAACATCAAAATTTGAGTGTGACCAAGCCCTTATGAGGCACTAAGGG 1938
QY 1966 CTATATCTTGAAGTTGTTGGAGATTAAGTACTAGTGAAGACCCCTGACGAAGC 2025
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1939 GTATATCTTGAAGATCTTTTGGCGACTGTCATGATGATGATGATGATGATGATGATG 1998
QY 2026 TCACATTTCTATGATGATGAGGAGCATGCAAGTGAAGCATGTTAGGCAAGAAAGCA 2085
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1999 TCCGATTCAGATATATGTTTGGTGAAGCTAGCAATGTTCAAAAGAAAACAAAATGG 2058
QY 2086 CAGTGAAGATCCAAATTAAGAACTTCAAGCGGATGAAGAGCCAGTGAAGAAACACCTGC 2145
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2059 TCAAGCTGATACCAATCAAGCGAAAAACCAAGCAGAGAAAACTCAGACAGAAAAAAC 2118
QY 2146 TGAGCCAGAAAGTCCCTCAAGTAGAGA 2171
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2119 TGAGGAAGAAAAACCTCGAAGAGA 2144
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 18
 ABQ84824
 ID ABQ84824 standard; DNA; 2290 BP.
 XX ABQ84824;
 AC
 XX
 DT 04-SEP-2002 (first entry)
 XX
 DE S. pneumoniae SP042 nucleotide sequence SEQ ID NO:65.
 XX
 KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
 antibacterial; Streptococcal infection; detection; gene; ds.

```

XX XX Streptococcus pneumoniae.
OS
XX US2002061545-A1.
PN
XX 23-MAY-2002.
PD
XX 22-JAN-2001; 2001US-0765272.
PF
XX 30-OCT-1997; 97US-0961083.
PR
XX
XX (CHOI/) CHOI G H.
PA (KUNS/) KUNSCH C A.
PA (BARA/) BARASH S C.
PA (DILL/) DILLON P J.
PA (DOUG/) DOUGHERTY B.
PA (FANN/) FANNON M R.
PA (ROSE/) ROSEN C A.
PI
PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;
PI Rosen CA;
XX WPI; 2002-479261/51.
XX P-PSDB; ABP54589.
XX
XX New Streptococcus pneumoniae antigens, useful for detecting
XX Streptococcus and for preventing or attenuating disease caused by
XX Streptococcus infection -
XX
XX Claim 1; Page 28-29; 70pp; English.
XX
XX AB084792 to AB084904 represents nucleic acids which encode the
XX Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.
XX The S. pneumoniae antigens have antibacterial activity and can be
XX used in vaccines. The S. pneumoniae antigens can also be used to
XX prevent or attenuate a Streptococcal infection in an animal. The
XX polynucleotides encoding the S. pneumoniae antigens can be used to
XX detect Streptococcus nucleic acids. AB084905 to AB085130 represent
XX primers used in the cloning of S. pneumoniae ORFs (open reading frames)
XX which are used in an example from the present invention.
XX
XX Sequence 2290 BP; 766 A; 474 C; 498 G; 547 T; 5 other;
SQ
Query Match 41.3%; Score 987.6; DB 24; Length 2290;
Best Local Similarity 67.7%; Pred. No. 5,7e-239;
Matches 1481; Conservative 0; Mismatches 645; Indels 60; Gaps 5;
QY 1 TTCTTACAGAGTTGGAGCTGTATCAAGCTAGAACGTTTAAAGAAA--TAACTGTGTTTC 57
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 4 TTCTTATGAAGCTTGCTGTGTCACCAAGCTGGTCAAGGTTAAGAAAGATCTAATGAGTTTC 63
QY 58 CTATATGATGAGAAAACAAGCGACGCAAAAAACGAGAAATTTGACTCTGATGAGGTTAG 117
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 64 TTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 123
QY 118 CAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTATCAAGTAAAGACAGCAAGGCTATGT 177
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 124 TAAAGGGAAGGAGATCAACGCCGACAAATGATGATGATGATGATGATGATGATGATGATG 183
QY 178 CACTTCAATGCGACCACTATCTATTTAATTAATGATGATGATGATGATGATGATGATGATG 237
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184 GACTCTTCATGAGAGACCATATATCTATTAATGCAAGGTTCTTATGATGATGATGATGATG 243
QY 238 CAGTGAAGAAATTAATCTGAAAGATCCAACTATTAAGCTAAAGATGAGATATGTTAA 297
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 244 CAGTGAAGAGCTCTCTATGAAAGATCCGAATTAATCAAGTGAAGATGATGATGATGATG 303
QY 298 TGAGTCAAGGAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 357
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 304 TGAATCAAGGAGTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 363
QY 358 TGCTGCCACGCGATTAAGCTCCGTACAAAGAGAAATCAATGACAAAGAAACAGAGCA 417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Db 364 TCCAGCTATGCGGATATATATTCGACAAAGAAGATTAAACGTACAGACGAGACG 423
Qy 418 TAGTCAACATCGTGAAGATGGAACCTCCAGAAAAGATGGCTGTTGCCCTGGACGTTCC 477
Db 424 CAGTCAATATCAT-----AACTCAAGAGAGATATATGCTGTTGCTGCAGCCGAGAC 474
Qy 478 GCAAGAGCGCTATCTACAGATGATGGTTATATCTTTAAATGCTTCGATATCATAGAGA 537
Db 475 CCAAGGAGCTTATACAAAGGATGATGGTATATCTTCATATGATCTGATATCATTTAGAGA 534
Qy 538 TACTGTGATGCTTATATGCTTCTCATGAGATCATTTACATTTACATTTCTTAAGATA 597
Db 535 CACGGGTATGCTTATATGCTTCTCATGAGATCATTTACATTTCTTAAGATA 594
Qy 598 GTTATCAGTACGAGTGGCTGCTGCAGAGGCTTCTATGCTGCGAGAAATCTGTC 657
Db 595 GTTATCAGTACGAGTGGCTGCTGCAGAGGCTTCTATGCTGCGAGAAATCTGTC 629
Qy 658 AAATTCAGAACCTTATGCGCGACAAAATAGCGATTAACATTCAGAACTGGGTACC 717
Db 630 -----ATTGGAATGGAGAGAGGATCTGCTCTTCTTCAAGTTTATATATATG 681
Qy 718 TTCTGTAGCAATCCAGAACTACAAATCTAACACAGACAAACAGACACTACAG 777
Db 682 AAATCCAGCTCAACCAAGATGTTCAGAGAACCAATCTGACTGTCTCAACTATCA 741
Qy 778 TCAAGCAATCAAGATATGACATGATGCTCTTGAAGAGCTTACAACTGCTT 837
Db 742 TCA-----AAATCAAGGGGAAAACATTTCAAGCTTTTACGTGAATGTATATGCTAAACCTT 798
Qy 838 GAGTCAACGACATGTGATCTGATGCGCTTGTCTTGTATCCAGACAAATCAACAGTGC 897
Db 799 ATCAGAGCGCATGTGGAATCTGATGCGCTTATTTTCGACCGGCAATCAACAGTGC 858
Qy 898 AACAGTGAAGTGTTCAGTGCACAGAGATCATTAACACTTATCTTACTTCA 957
Db 859 AACCGGCAAGGTGATGCTGCTCTCATGTATCACTTATCACTTATCTTATTAACA 918
Qy 958 AATGCTGAATGGAAGACGAATGCTGATTTATCCCTTGTGATTCGTTCAACCA 1017
Db 919 AATGCTGAATGGAAGACGAATGCTGATTTATCCCTTGTGATTCGTTCAACCA 978
Qy 1018 TTGGGTACGAGATTCAGAGCCAGAACCAACCAAGTCCACACAGCTCCGGAACCTAGTCC 1077
Db 979 TTGGGTACGAGATTCAGAGCCAGAACCAACCAAGTCCACACAGCTCCGGAACCTAGTCC 1038
Qy 1078 AGCGCCGCAACTGCACCAATCTTAAATATGACTCAAAATCTTC-----TTT 1125
Db 1039 AAGTCGCAACTGCACCAATCTTCACACAGCTCCACCAATCTTCATGATGAGAAAT 1098
Qy 1126 GGTATGCTGCTGATCGAAGAAAGTGGGAGAGATGTATTCGAGAAAGAGGCTCTC 1185
Db 1099 GGTCAAGAGAGTGTTCGAAAGTAGGCGATGTATGCTTTCGAGAAAGAGGCTTTC 1158
Qy 1186 TCGTATGCTTTCGAGAAAGTTCACATCTGAAACTGTTAAAAATCTTGAAGCAAGTT 1245
Db 1159 TCGTATATCCCAAGCAAGATCTTTCACAGAAAAGACAGAGATGTATGACCAACT 1218
Qy 1246 ATCAAAAACAGAGAGTTCACACACTTAACTGCTAAAGAAAAGAAATGTTGCTCTCG 1305
Db 1219 GGCCTAAGCAGAAAGTTATCTCATAGCTAGAGCTAAGAAAACCTCCCATCTAG 1278
Qy 1306 TGACCAAGAAATTTATGATTAAGCAATATCTGTTAACTGAGGCTCATAAAGCCTTGT 1365
Db 1279 TGATGAGAAATTTTCAATTAAGCTTATGACTTACAGAGAAATTCACCAAGATTTACT 1338
Qy 1366 TGNAAATAGAGGCTGTAATCTGATTTCCAGGCTTACAGCAATTTATGAAGGCTTGA 1425
Db 1339 TGATATATTAAGTCCAGCAAGTGTATTTGAGGCTTGGATTAACCTTTGGAAGCACTCA 1398
Qy 1426 TGATGAATCGACTAATAAAGAAAATTTGATAGATGATTTATTTGCAATTCCTAGCAAT 1485
Db 1399 GATGTCTCNCATGATTAAGTCAAGTTAGTGANATATTTCTTGCTTCTTATGCTCGAT 1458

Qy 1486 TACCATCCAGAGCAGCTTGGCAAAACCAATTTCTAAATTTAGATATCTAGAACGAACT 1545
Db 1459 TCGATATCCAGAACCTTTAGGAAAACCAATTCGAAATTTACTTACATGATGATGAGT 1518
Qy 1546 TCGATTTGCTCAATTAGCTGATTAAGTATACAACTGATGTTTACATTTTATGATGA 1605
Db 1519 TCAAGTAGCAAGTTGGCAGGCAAGTATACAAACAGAAAGCGTTATATCTTTGATCTCG 1578
Qy 1606 TGATTAATCAGTATGAAGAGATGATATGTAACGCTCATATGAGGCTATGATCTG 1665
Db 1579 TGATTAATCAGTATGAGGAGGATGCTATGTAACCTCAATATGACCAATAGCACATG 1638
Qy 1666 GATTGAAAAGTACCTTTTGTATTAAGAAAAGTGCAGCTCAAGCTTATATTAAGA 1725
Db 1639 GATTAAAAAGTATGTTGTCTGAAGCTGAGAGAGCGGCGCAGGCTTATGCTTAAGA 1698
Qy 1726 AAAAGTATCTTACCTCATCTCCAGACGACAGATTTAAAGCAATCCAACTGAGATAG 1785
Db 1699 GAAAGTTTGAACCTCTCTTGCAGACATCAGATTCAGGAATATCAGAGCAAAAG 1758
Qy 1786 TGCACAGCTATTTCAATCTGTGAAGGGAAGAAAGCAATTCACCTGTTGACTTC 1845
Db 1759 AGCAGAGCTATCTCAACCGGTGAAGCAAGCTTAAGAGGTGCCAATGATCGTATGCC 1818
Qy 1846 ATATATGTTGACATCAGTGAAGTTAAAAAGGTAATTTGATTTATCTCATTAAGA 1905
Db 1819 TTCAATCTCAATATACATGTAAGTAAAAAGGTAATTTATATATCTCATTAATA 1878
Qy 1906 TCATTAACATTAATTAATTTGCTTGTGATGATCACATCAAAAGCTCCAAATAG 1965
Db 1879 CCATTAACATTAATTAATTTGATGATGATGATGATGATGATGATGATGATGATGATG 1938
Qy 1966 CTATATCTTGAAGATTTGTTGCGAGATTAAGTATGATGATGATGATGATGATGATG 2025
Db 1939 GTATATCTTGAAGATTTGTTGCGAGATTAAGTATGATGATGATGATGATGATGATG 1998
Qy 2026 TCCACATCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2085
Db 1999 TCCACATCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2058
Qy 2086 CAGTGAAGTCCAAATTAAGAACTTCAAGCGGATGAAGCGGATGAAGCAAGTGAAG 2145
Db 2059 TCAAGCTGATTAATTAATTAAGAACTTCAAGCGGATGAAGCGGATGAAGCAAGTGAAG 2118
Qy 2146 TGAGCCAGAAAGTCTCTCAAGTAGAGA 2171
Db 2119 TGAGGAGAAAGCCCTCGAGAAAGAGA 2144

RESULT 19
AAA65737
ID AAA65737 standard; DNA; 2639 BP.
XX
XX
XX
XX
DT 21-NOV-2000 (first entry)
XX
XX
XX Streptococcus pneumoniae BVH-11-2 gene SEQ ID NO:13.
XX
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal; ds.
XX
XX Streptococcus pneumoniae.
XX
XX
XX PN- WO200039299-A2.
XX PD 06-JUL-2000.
XX
XX PF 20-DEC-1999; 99WO-CA01218.
XX
XX PR 23-DEC-1998; 98US-0113800.

XX (BIOC-) BIOCHEM PHARMA INC.
 PA Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX WPI: 2000-453297/39.
 XX P-PSDB; AAB12720.
 XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteremia and/or pneumonia -
 XX Example 6; Fig 16; 106pp; English.
 XX The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence encodes the
 CC S. pneumoniae BvH-11-2 protein antigen.
 XX
 SQ Sequence 2639 BP; 889 A; 518 C; 567 G; 665 T; 0 other;
 Query Match 41.1%; Score 980.8; DB 21; Length 2639;
 Best Local Similarity 67.8%; Pred. No. 3.2e-237;
 Matches 1473; Conservative 0; Mismatches 638; Indels 60; Gaps 5;

1 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAACGTTAAGAAAA--TAATCGTGTTC 57
 Db 173 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAACGTTAAGAAAAAGTCTAATCGAGTTTC 232
 Qy 58 CTATTAATAGGAAACAGACGACGCAAAACGGAATTTGACTCCTGATGAGGTAG 117
 Db 233 TTATATATAGTGTGTACAGGCTGTGTCAAAAGGCAAAATTTGACACAGATGAAGTAG 292
 Qy 118 CAAGCGTGAAGAAATGCTGAGCAAAATGCTCAAGATTAACAGACCAAGCTATGT 177
 Db 293 TAAAGAGAGGGGATCAACGCGCAAAATTTGTTCAAGATTAACAGATTAAGTTAGT 352
 Qy 178 CACTTACATGAGCGACACTATCATTTATTAACATGTAAGGTTCTTATGAAGCTATCAT 237
 Db 353 GACCTCATGAGACATTAATCATTTATGGAAGGTTCTTATGATGACCATCAT 412
 Qy 238 CAGTGAATAATTTACTGATGAAGAATCCAACTATTAAGCTAAAGATGAGATTTGTTAA 297
 Db 413 CAGTGAATAATTTACTGATGAAGAATCCAACTATTAAGCTAAAGATGAGATTTGTTAA 472
 Qy 298 TGAGGTCAAGGTTGATATGTTATCAAGGTAGATGGAATAATCTATGTTTACCTTAAGA 357
 Db 473 TGAATTAAGGTTGCTATGTTATTAAGGTAGACGGAATAATCTATGTTTACCTTAAGA 532
 Qy 358 TGCTGCCACGCGGATAACGTCCTGATCAAAAGAGAAATCAATGACAAAAACAAGACA 417
 Db 533 TGCGGCCCATGCGGACATTAATTCGGAACAAAGAAAGATTAAACGTGAGAGACAGACA 592
 Qy 418 TAGTCAATCGTGAAGGTGAATCCCAAGAAAGATGCTGTTGCTTGGCCTTGACGCTTC 477
 Db 593 CAGTCAATATCAT-----AACTCAAGAGCAATATGCTGTTGCTGACGCGACAGC 643
 Qy 478 GCAAGAGCGTATATCTACAGATGATGTTATCTTTATGCTTATGATCATATCATAGAGA 537
 Db 644 CCAAGAGCGTATATCAAGATGATGTTATCTTTATGCTTATGATCATATCATAGAGA 703
 Qy 538 TACTGTGATGCTTATATGTTCTCTCATGAGATCATTAACATTAATCTCTTAAGATGA 597
 Db 704 CACGGGTATGCTTATATGTTCTCTCATGAGATCATTAACATTAATCTCTTAAGATGA 763
 Qy 598 GTTATCACTGACGAGTTGGCTGCTGCAAGACCTTCTATCTGTGAGAGAAATCTGTC 657
 Db 764 GTTATCACTGACGAGTTGGCTGCTGCAAGACCTTCTATCTGTGAGAGAAATCTGTC 798

Qy 658 AAATTCAGAACCTATGCGCCACAAAATAGCATATACATTCAAGAACAACTGGGTACC 717
 Db 799 -----ATTGAATGGGAAGGAGATCTGCTTCTTCAAGTTCTAGTTATATGTC 850
 Qy 718 TTCTGTAGCAATTCAGAGACTACAAATATCTAACAGACAAACAGCAACTAATACAG 777
 Db 851 AAATCCAGTTCAACCAAGATGTGAGAGAACCAACATCTGCTGCTCACTCCACTTATCA 910
 Qy 778 TCAAGCAAGTCAAGTATATGATGATGATGCTCTTGAAGACGCTTACAAATGCGCTTT 837
 Db 911 TCA---AAATCAAGGGGAAAAATTCAAGCCTTTTACGATATGTTATGCTAAACCTT 967
 Qy 838 GAGTCAACGACATGATAGATGATGCTGCTTGTCTTATGATCAGACCAATCACAAGTCG 897
 Db 968 ATCAGAACGCGATGATGATGATGCTGCTTATGATGATGATGATGATGATGATGATG 1027
 Qy 898 AACGCTAGAGGTGTGACGCTGACACAGAGATCATTTACACTTCACTTCTTACTCTCA 957
 Db 1028 AACCGCAGAGTGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1087
 Qy 958 AATGCTGAATTTGAGAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1017
 Db 1088 AATGCTGAATTTGAGAGAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1147
 Qy 1018 TTGGGTACAGATTTCAAGGCGACAAACCAAGTCAACACGACTCCGAACTAGTCC 1077
 Db 1148 TTGGGTACAGATTTCAAGGCGACAAACCAAGTCAACACGACTCCGAACTAGTCC 1207
 Qy 1078 AGCCCGCAACCTGACCACTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTTC 1125
 Db 1208 AAGCTGCAACCTGACCACTTCAATTTCAATTTCAATTTCAATTTCAATTTCAATTTTC 1267
 Qy 1126 GGTATGAGCTGTGTAAGAAAGTTGGGGAAGATGATGATGATGATGATGATGATGATG 1185
 Db 1268 GGTATGAGCTGTGTAAGAAAGTTGGGGAAGATGATGATGATGATGATGATGATGATG 1327
 Qy 1186 TCGTTATGCTTGTGCGAAGATTTTACCATGTAATCTGTAATCTGTAATCTGTAATCTG 1245
 Db 1328 TCGTTATGCTTGTGCGAAGATTTTACCATGTAATCTGTAATCTGTAATCTGTAATCTG 1387
 Qy 1246 ATCAAAACAGAGAGTTTTCACACACTTATGCTTATGCTTATGCTTATGCTTATGCT 1305
 Db 1388 GGCACACAGAGAGTTTATCTCATAGCTAGAGCTAGAGAACTGACCTCCATCTAG 1447
 Qy 1306 TGACCAAGATTTTATATAGATTAAGCATTAATCTGTTAAGTGAAGCTGATTAAGCTGTT 1365
 Db 1448 TGATCGAAGATTTTATATAGATTAAGCATTAATCTGTTAAGTGAAGCTGATTAAGCTG 1507
 Qy 1366 TGNAAATAGAGTGTGTAATCTGATTTCAAGCTTGAACCAATTTATAGAACGCTTGA 1425
 Db 1508 TGATTAATAGAGTGTGTAATCTGATTTCAAGCTTGAACCAATTTATAGAACGCTTGA 1567
 Qy 1426 TGATGAATCGATTAATTAAGAAAAATTTGATGATGATTTATGATGATTTCTAGCAAT 1485
 Db 1568 GGATGCTCAAGTATTAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1627
 Qy 1486 TAACCATCGAGAGGACCTGAGCAAAACCAATTTCAATGAGATTAAGTGAAGCAAGT 1545
 Db 1628 TCGTATTCGAGAGGACCTGAGCAAAACCAATTTCAATGAGATTAAGTGAAGCAAGT 1687
 Qy 1546 TCGTATTCGAGAGGACCTGAGCAAAACCAATTTCAATGAGATTAAGTGAAGCAAGT 1605
 Db 1688 TCAAGTACGCAAGTTGCGACAGCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1747
 Qy 1606 TGATTAATCAAGTGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1665
 Db 1748 TGATTAATCAAGTGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1807
 Qy 1666 GATTGAAGAGATGAGCTTCTGTAAGAAAAATTTGAGCAAGTGAAGCAAGTGAAGCAAG 1725
 Db 1808 GATTGAAGAGATGAGCTTCTGTAAGAAAAATTTGAGCAAGTGAAGCAAGTGAAGCAAG 1867
 Qy 1726 AAAAGTATCTTACCTCATCTCCAGACGAGATGTTAAAGCAATCAACTGAGATAG 1785

```

Db      1868  GAAAGGTTGACCCCTCCCTTCAGACAGACCAGGATTCAGAAATATGAGGCAAAAGG 1927
Oy      1786  TGCAGACGATTTTCAATCGGTGAAGGGGAAAACGAATTCACCTGTTCCGACTTC 1845
Db      1928  AGCAGAACCTCTACCAACCGGTGAAGCACTAAGAGGTCACCTGATCGTATGCC 1987
Oy      1846  ATATATGTTGAGCATACAGTTGAGGTTAAACGGTAATTGATTATTCCTCATAGGA 1905
Db      1988  TTACATCTTCAATATCTGTAAGTCAAAAACGGTAGTTAATCATACCTCATTAAGA 2047
Oy      1906  TCATTACCAATATATTAATTTGCTGTTGATGATCAACATACAAAGCTCCAAATGG 1965
Db      2048  CCATTACCAATACATCAAAATTTGAGTGTGACCAAGGCTTTATGAGGCACCTAAGG 2107
Oy      1966  CTATACCTTGAAGATTTGTTTGGACGATTAAGTACTACGATGAAACCCCTGAGGAG 2025
Db      2108  GTATATGCTTGGAGATCTTTGGCCAGCTGTCAAGTACTATGTGCAACATCCAAACG 2167
Oy      2026  TCCACATCTAATGATGATGGGCAATGCCAGTACGATGTTAGGCAAGAAAGACCA 2085
Db      2168  TCCGATTCAGATATGTTGTTGGTACGCTAGTACCATGTTGTAATAAATTAAGGACA 2227
Oy      2086  CAGTGAAGATCCAAATTAAGAACTTCAAGCGGATGAAGCCAGTAGAGGAAACACCTGC 2145
Db      2228  CCAAGATAGTAACTGATGAAGATAGGACATGATGAAGTGAAGCCAACTCACC 2287
Oy      2146  TGAGCCAGAG 2156
Db      2288  TGAATCTGATG 2298

```

RESULT 20

```

ABK15104
ID      ABK15104 standard; DNA; 2639 BP.
XX      ABK15104;
XX      DT      08-MAY-2002 (first entry)
XX      DE      DNA encoding Streptococcus pneumoniae BVH-11-2.
XX      KW      BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX      KW      pneumonia; streptococcal bacterial infection; gene; ds; BVH-11-2.
XX      OS      Streptococcus pneumoniae.
XX      FH      Key
XX      FT      CDS
XX      FT      Location/Qualifiers
XX      FT      114..2630
XX      FT      /tag= a
XX      FT      /product= "BVH-11-2"
XX      FT      /note= "The gene is flanked by sequences from the
XX      FT      vector SP64, no information on which is
XX      FT      given in the specification"
XX      PN      MO200198334-A2.
XX      PD      27-DEC-2001.
XX      PF      19-JUN-2001; 2001WO-CA00908.
XX      PR      20-JUN-2000; 2000US-212683P.
XX      PA      (SHIR-) SHIRE BIOCHEM INC.
XX      PI      Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX      DR      WPI; 2002-122272/16.
XX      DR      P-PSDB; AAU75934.
XX      PT      New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX      PT      epitope-bearing polypeptides, useful as vaccine components for treating
XX      PT      or preventing streptococcal infections such as otitis media,

```

```

PT      meningitis, and bacteraemia
XX
XX      Example 3; Fig 5; 113pp; English.
PS
XX      CC      The invention describes an isolated polypeptide (I) with 70-90%
XX      CC      identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX      CC      BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX      CC      comprising (I) is useful for therapeutic or prophylactic treatment of
XX      CC      meningitis, otitis media, bacteraemia or pneumonia infection in an
XX      CC      individual susceptible to these disorders. (II) is also useful for
XX      CC      therapeutic or prophylactic treatment of any streptococcal bacterial
XX      CC      infection (e.g., caused by Streptococcus pneumoniae, group A
XX      CC      Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
XX      CC      as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novardi or
XX      CC      Staphylococcus aureus) in an individual susceptible to the infection.
XX      CC      A polynucleotide (III) encoding (I) is useful in DNA immunisation
XX      CC      techniques. The Streptococcus polypeptides are useful in a diagnostic
XX      CC      test for S. pneumoniae infection. (III) is useful for designing DNA
XX      CC      probes for use in detecting the presence of Streptococcus in a biological
XX      CC      sample suspected of containing the bacteria. The DNA probes may also be
XX      CC      used for detecting circulating S. pneumonia nucleic acid in a sample for
XX      CC      diagnosing streptococcal infections. This sequence encodes the
XX      CC      Streptococcus pneumonia protein BVH-11-2, used to create the antigenic
XX      CC      peptides described in the method of the invention.
SQ      Sequence 2639 BP; 889 A; 518 C; 567 G; 665 T; 0 other;

```

```

Query Match      41.1%; Score 980.8; DB 24; Length 2639;
Best Local Similarity 67.8%; Pred. No. 3.2e-237;
Matches 1473; Conservative 0; Mismatches 638; Indels 60; Gaps 5;

```

```

Oy      1  TTCTTACGATTTGGAGCTGTATCAAGCTAAGGTTAGGAAA---TAATCGGTTTC 57
Db      173  TTCTTATACACTGTGCTGCTACCAAGCTGTCAGGTTAAGAAAGTCTAATCGAGTTTC 232
Oy      58  CTATATAGATGAAAAACAAGCGAGCGCAAAAACGAGATTTGACTCTGATGAGGTTAG 117
Db      233  TTATATATAGATGATATCAGCTGTGTCBAAGGCGAANAATTTGACACCAAGATGAGTCAG 292
Oy      118  CAAGCTGAAGAAATCAATGCTGAGCAAAATGCTCATCAAGATTAACAGCAACAGGCTATGT 177
Db      293  TAAAGAGAGAGGGGATCAACGCCGAACAATTTGTTACAAATTAACGATCAAGGTTATGT 352
Oy      178  CACTTCAATAGGCAACCACTATCAATTTAACAAGTGAAGTTCCTTATGACGATATCAT 237
Db      353  GACCTCTATGAGGCCATTAATCACTACTAATTAATGCAAGGTTCTTATGATGTCATAT 412
Oy      238  CAGTGAAGATTAATCATGAAAGATCCAAACTATTAAGCTAATAAAGTGAAGATATTGTTAA 297
Db      413  CAGTGAAGAACTTCTCATGAAAGATCCGAATTTATCAGTTGAAGATTCAGACATTGTCAA 472
Oy      298  TGAGGTCAAGGGTGAATATGTTATCAAGGTAGATGAAGAAATTAATGTTTAACTTAAGA 357
Db      473  TGAATCAAGGGTGGCTATGATTAAGTTAGACGGAANAATATCATGTTTAACTTAAGA 532
Oy      358  TGTGTCACACCGCGATTAACGTCCTGTAACAAGAGAAATCAATGCAAAAAACAAGAGCA 417
Db      533  TCGGCGCCATCGGCAATATTTCCGACAAAAGAAAGATTAACCTGCAAGACAGAGAAACA 592
Oy      418  TAGTCAACATCGTGAAGGTGAACCTCAAGAAACGATGCTGTTGCTTGGACGCTTC 477
Db      593  CAGTATATATAT-----AACTCAAGACAGATATATGCTGTTGCTGACGACGAGAGC 643
Oy      478  GCAAGAGCGTAACTACAGATGATGTTATATCTTTAATGCTTGTGATATCATAGAGA 537
Db      644  CCAAGAGCGTTATACAAAGGATGATGGGTATATCTTCAATCATCATGATATCATATGAGA 703
Oy      538  TACTGTGATGCTTATATGCTTCTTCAAGAGATATTAACATTAATTTCTTAAGAAAGA 597
Db      704  CACGGTATGCTTATATGCTTCTTCAAGGAGACCATTAATTAATTTCTTAAGAAAGA 763
Oy      598  GTTATACAGTACGAGTTGGCTGTCAGAGACCTTCTTATCTGATCGAGAAATCTGTC 657

```

```

Db      764 GTATACAGTACGAGTGTAGCTGTGAGAAAGCT----- 798
Oy      658 AAATTCAAGAACCTATGCGCCGCAAAATAGGATTAACCTTCAAGAACAACTGGGCTAC 717
Db      799 -----ATTGGAATGGGAAGCAGGGAATCTGCTCTTCTTCAAGTTCTTATATATGC 850
Oy      718 TTCTGTAAAGCAATCAGAACTACAAATATCTAACAGAACAAACAGCAACTAACG 777
Db      851 AAATCACTTCAACCAAGATTGTGAGAACCAAACTGACTGTCACTCCAACTATCA 910
Oy      778 TCAGCAAGTCAAGTAATGACATTGATAGTCTTTGAAACAGCTTCAAACTGCTTT 837
Db      911 TCA---AAATCAAGGGGAAAAATTTCAGGCTTTTACGTGATGTGCTAAACCTT 967
Oy      838 GAGTCAAGGACATGATGATGATGAGGCTGTCTTATCCAGACAAATCACAAGTCG 897
Db      968 ATCAGAACGCCATGATGATGATGAGGCTTATTTTCAACCGAGGCAAAATCACAAGTCG 1027
Oy      898 AACAGCTAGAGTGTGGAGTGCACACAGGATCATTAACACTTCACTCCCTACTCA 957
Db      1028 AACGCCAGAGGTGATGCTGCTCCCTCATGTAACATTACACTTATCCCTATGAGACA 1087
Oy      958 AATGCTGAATTTGAGAACGATGCTGCTATTTATCCCTGCTTATGCTTCAACA 1017
Db      1088 AATGCTGAATTTGAGAAAAAGATGCTGTAATATCCCTTCTTATGCTTCAACA 1147
Oy      1018 TTGGGTACAGATTTCAAGGCCGAGAACCAAGTCCACACCGACTCCGGAACCTAGTC 1077
Db      1148 TTGGGTACAGATTTCAAGAACCAAGTCCACACCGACTCCGGAACCTAGTC 1207
Oy      1078 AGGCCCGCAACCTGCACCAATCTTAAATAGACTCAAAATCTTC-----TTT 1125
Db      1208 AAGTGTGCACTGTCACCAAAATCTCACAACAGCTCCAAAGCAATCAATGATGAGAAAT 1267
Oy      1126 GGTATGACGCTGTGACAAAAGTTGGGAGAGATGATGATTCGAGAAAAGGCTATC 1185
Db      1268 GGTCAAAAGAGCTGTTGAAAAGTAGCGGATGATGATGTTGAGGAAATGAGTTTC 1327
Oy      1186 TCGTATGCTCTTTGGGAAGATTTACATCTGAACCTGTAAATCTTGAAGCAAT 1245
Db      1328 TCGTATATCCAGGCAAGGATCTTTGACAGAAAACACAGGCAATGATGAGCAACT 1387
Oy      1246 ATCAAAAGAGAGATGTTTCAACACTTAACTCTAAAAAGAAAATGTTGCTCTCG 1305
Db      1388 GGCACAGCAGAAATTTATCTCAATAGTACGACTAAGAAAATGACCTCCATCAG 1447
Oy      1306 TGACCAAGATTTATGATTAAGATATATCTGTTAACTGAGCTCATAAAGCTTGT 1365
Db      1448 TGATCGAAGATTTTCAATTAAGGCTTATGACTTACAGAAATTCACCAAGATTACT 1507
Oy      1366 TGNAAATAGGCTCGTAATTCGATTTCCAGCCTTAACAAATATTAAGAACCTTGA 1425
Db      1508 TGATATTAAGGTGCAAGTTGATTTTGAAGTTCGTTGAAACGACTCA 1567
Oy      1426 TGATGATCGACTAATTAAGAAAATTTGATAGATTTATGAGCATTCCTAGCACT 1485
Db      1568 GGAATCTCAAGTATTAAGTCAAGTATGATATTTCTGCTTCTTAAGCTCCAT 1627
Oy      1486 TACCATTCAGAGGACTTGGCAACCAATTTCAAAATGATGATGATGAGCAAGT 1545
Db      1628 TCGTATCCAGAACCTTTAGGAAAAACCAATGCGCAAAATTAACCTACATGATGAGAT 1687
Oy      1546 TCGTATGCTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1605
Db      1688 TCAAGTACCAAGTTGGCAGGCAAGTACCAAGCAAGCGGTATATCTTGTGATCTCG 1747
Oy      1606 TGATATTAAGTATGAGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1665
Db      1748 TGATATTAAGTATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1807
Oy      1666 GATTGAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1725
Db      1808 GATTGAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1867

```

```

Oy      1726 AAAAGTATCCTACCTCCATCTCCAGACGAGATGTTAAAGCAATCCAATCGAGATAG 1785
Db      1868 GAAAGTTTACCCCTCTCTTCAGACACACAGATTCAGAAATCTGAGGAAAAAG 1927
Oy      1786 TGACAGCTATTTTCAATCGTGTGAAAAGGAAAAAGAAATTCACCTGTTGACTTCC 1845
Db      1928 AGCAGAGCTATCTACAAACCGCGTGAAGCAGTAAAGGTGCTCATGATGCTATGCC 1987
Oy      1846 ATATATGTTGAGATCAGTATGAGTTAAACCGTAATTTGATTTATCTCTCAATAGA 1905
Db      1988 TTCAATCTTCAATATACGTAGAAATCAAAAACGTAAGTTATATCACTCATATAGA 2047
Oy      1906 TCATTACCAATATATTAATTTGCTTGTGATGATCACAATACAAAGCTCCAAATGG 1965
Db      2048 CCATTACCAATATCAATTAATTTGCTTGTGATGATGATGATGATGATGATGATGAT 2107
Oy      1966 CTATACCTTGGAAGATTTGTTTGGCAGATTAAGTACTAGTACACACCTGAGAACG 2025
Db      2108 GTATAGCTTGAAGATCTTTTGGCAGCTGTCAGATGATGATGATGATGATGATGAT 2167
Oy      2026 TCCACATTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2085
Db      2168 TCCGATTCAGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2227
Oy      2086 CAGTAAAGATCCAAATTAAGAACTTCAAGCGGATGAAGACAGTAAAGCAACCTGTC 2145
Db      2228 CCAAGATGATGAACCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 2287
Oy      2146 TGAGCCAGAAAG 2156
Db      2288 TGAATCTGATG 2298

```

```

RESULT 21
AAK25394
ID AAK25394 standard; DNA; 2163 BP.
XX
AC AAK25394;
XX
DT 19-JUL-1999 (first entry)
XX
DE Streptococcus pneumoniae complement C3-degrading protease DNA.
XX
KW Human complement C3-degrading protease; vaccine; infection;
KW meningitis; pneumonia; xeroderma; transplantation; transplant rejection;
KW inflammation; ds.
XX
OS Streptococcus pneumoniae.
XX
PN WO915675-A1.
XX
PD 01-APR-1999.
XX
PF 24-SEP-1998; 98WO-US20186.
XX
PR 24-SEP-1997; 97US-0059907.
XX
PA (AMCY ) AMERICAN CYANAMID CO.
XX
PA (MINU ) UNIV MINNESOTA.
XX
PI Cheng Q, Finkel DJ, Green BA, Hostetler MK, Masi AW;
DR WPI; 1999-254719/21.
DR P-PSDB; AAY05753.
XX
PT New isolated human complement C3-degrading proteinase
XX
PS Claim 54; Page 52-54; 66pp; English.
XX
CC This DNA sequence encodes a 79 kDa protein (see AAY05753) of
CC Streptococcus pneumoniae serotype 4 that is capable of degrading
CC human complement protein C3 (HCPc3). It was identified in the S.

```

CC pneumoniae serotype 4 genome by alignment to another novel open
 CC reading frame (see AAX25393) that codes for a 20 kDa HPC3 protease
 CC (AAY05152) of *S. pneumoniae* serotype 3. This suggested the open
 CC reading frame that codes for the 20 kDa protein may be part of a
 CC larger open reading frame. Amino acids 1-58 and 90-133 of the 20
 CC kDa protein have substantial sequence identity with amino acids
 CC 170-227 and 258-300 of the 79 kDa protein. Proteins and peptides or
 CC polypeptides containing these regions, and DNA sequences encoding
 CC them (nucleotides 507-681 and 827-999 of the present sequence) are
 CC claimed. HPC3 proteases and polypeptides can be used as immune
 CC system stimulating compositions (claimed). They can produce an
 CC immune response against *S. pneumoniae* to immunize or treat a
 CC mammalian subject against infection or colonization (claimed).
 CC They can produce a B cell response, a T cell response, an
 CC epithelial cell response, or an endothelial cell response
 CC (claimed). The expression of the proteins on the surface of an
 CC organ of an animal used in xenotransplantation can be used to
 CC inhibit C3-mediated inflammation and rejection.

Sequence 2163 BP; 737 A; 450 C; 462 G; 514 T; 0 other;

Query Match 36.4%; Score 869.4; DB 20; Length 2163;

Best Local Similarity 65.2%; Pred. No. 3.9e-209;

Matches 1436; Conservative 0; Mismatches 627; Indels 141; Gaps 5;

254 ATGAAAGATCCAACTATAAGCTAAAGATGAGATATTGTTAATGAGTCAAGGTGGA 313
 1 ATGAAAGATCCAACTATAAGCTATAAGCTATAAGCTATAAGCTATAAGCTATAAG 60
 314 TATGTTATCAAGTATGATGAAATATCTATGTTTACCTTAAGAGTCCGACGGGAT 373
 61 TATGTTATCAAGTATGATGAAATATCTATGTTTACCTTAAGAGTCCGACGGGAT 120
 374 AACGTCGATCAAAAGAGAAATCAATCGACAAAACAGAGCATGTCAATGCTGAA 433
 121 AATATTCGACAAAAGAGAAATCAATCGACAAAACAGAGCATGTCAATGCTGAA 180
 434 GGTGAACTCCAGAAAGCATGATGCTGCTTGGCCTTGACAGCTTGGCAAGAGCATTA 493
 181 GGTGATCT 234
 494 ACAGATGATGTTATATCTTAAATGCTTGAATATCAAGAGATATCTGATGCTTAT 553
 235 ACGGATGATGTTATATCTTAAATGCTTGAATATCAAGAGATATCTGATGCTTAT 294
 554 ATCGTTCTCATGAGATATTAACCTTATCTTGAATATCAAGAGATATCTGATGCTTAT 613
 295 ATCGTTCTCATGAGATATTAACCTTATCTTGAATATCAAGAGATATCTGATGCTTAT 354
 614 TTGGCTGCTGAGAAAGCTTCTATCTGATGCTGAGAAATCTGCAAAATTCAGAACTTAT 673
 355 TTAGCTGCTGAGAAAGCTTCTATCTGATGCTGAGAAATCTGCAAAATTCAGAACTTAT 381
 674 CCACGCAAAATATGAGATATCACTTCAAGAACTGAGTACTTCTGTAAGCAATCCA 733
 382 GGGAGACAGGAGTCTGCTCTTCTTCAATGTTAGTTATTAATGCAATCCAGCTACCA 441
 734 GGAATCAATATCTTCAACAAAGCAACAAAGCACTTACAGTCAAGCAAGTCAAGT 793
 442 AGATTGTCAGAGAACCAATCTGATCTGATCTCAACTTCACTTCACTTCACTTCACT 498
 794 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 853
 499 GAAATATTTTCAAGCTTTTACGTTAGTTATGTTAATCCCTTATCAGAACGCACTG 558
 854 GAATCTGATGAGCTTGTCTTGTATCCAGCAAAATCAAGTCAAGTCAAGTCAAGTCA 913
 559 GAATCTGATGAGCTTGTCTTGTATCCAGCAAAATCAAGTCAAGTCAAGTCAAGTCA 618
 914 GCAATGCAACAGAGATATTAACCTTATCTTGAATATCAAGTCAAGTCAAGTCAAGT 973
 619 GCTGTCCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 678

974 GAACGAATGCTGCTGATTAATTCCTCTGCTTATCTGTCMAAACATGGGTACAGATTCA 1033
 679 AAACGAATTTGCTGCTGATTAATTCCTCTGCTTATCTGTCMAAACATGGGTACAGATTCA 738
 1034 AGGCCAGAACCAACAGTCCACACCGACTCCGGAACCTAGTCCAGGCCGCACTGCA 1093
 739 AGACGAGAACCAACAGTCCACACCGACTCCGGAACCTAGTCCAGGCCGCACTGCA 798
 1094 CCAATCTTAAATATGATGCA-----ATTCTTCTTGGTATGATGATGATGATGATGAT 1141
 799 CCAATCTTAAATATGATGCA-----ATTCTTCTTGGTATGATGATGATGATGATGATGAT 858
 1142 CGAAAGTGGGGAAGATATGATATTCGAGAAAGAGGCACTCTGTTATGCTTTGGC 1201
 859 CGAAAGTGGGGAAGATATGATATTCGAGAAAGAGGCACTCTGTTATGCTTTGGC 918
 1202 AAAAGTTTACCATCTGAAATCTGTTAAATCTTGAAGCAAGTTATCAAAACAGAGAGT 1261
 919 AAGGATCTTTCAGCAGAAACAGACGAGCATGTTATCAAACTGGCCAGAGAAAGT 978
 1262 GTTTCACACCTTTTAACTGCTTAAAGAAATGTTGCTCTCTGACCAAGATTTTAT 1321
 979 TTATCTTAAAGCTTGAAGCTTAAAGAACTGACCTCCATCTTAAAGTGAAGATTTTAC 1038
 1322 GATAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTGTTGTTGTTGTTGTTGTT 1381
 1039 AATAAGCTTATGATCTTACTAGCAAGATTTACCAAGATTTTACTGTTATTAAGAGTCA 1098
 1382 AATTCGATTTTCAAGCTTACAGCAAAATTTATGAAAGCCTTGAATGATGATGATGAT 1441
 1099 CAAGTTGATTTTGAAGCTTTGATTAACCTGTTGAAGCACTGAGATGATGATGATGAT 1158
 1442 AAAAGAAATTTGATGATGATTTATTTGATTTCTTGAAGCACTGAGATGATGATGATGAT 1501
 1159 AAAAGCAATTTGATGATGATTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1218
 1502 CTTGCAAAACCAAAATTTCTCAATTTAGTATCTGTAAGCAAGATTTCTGATTTCTCAATTA 1561
 1219 TTAGGAAACCAAAATTTCTCAATTTAGTATCTGTAAGCAAGATTTCTGATTTCTCAATTA 1278
 1562 GCTGATTAAGTATCAACGTCAGATGTTATCATTTTGTATGAAACATGATTAATCAATGAT 1621
 1279 GCAGGCAAGTATCAACGTCAGATGTTATCATTTTGTATGAAACATGATTAATCAATGAT 1338
 1622 GAAAGATGATATGATTAAGCTTATTTGAGGCTTATGAGGCTTATGAGGCTTATGAGGCT 1681
 1339 GAGGGGATGCTTATGATTAATCTTCAATTTGATGATGATGATGATGATGATGATGATGAT 1398
 1682 CTTTCTGATTAAGGAAAGTTGACGCTCAAGCTTATTAAGGAAAGGATTTCTGATCT 1741
 1399 TTGCTTGAAGCTGAGAGAGGCGGACCGAGCTTATGCTTAAAGAAAGGTTTGACCTCT 1458
 1742 CCATCTCAGACGAGATGTTTAAAGCAATTCMACTGAGATGTCAGAGCTATTTAC 1801
 1459 CTTGCAACAGCATCAGAGATTCAGAAATCTAGAGGAAAGGAGGAGGATTTGACCTCT 1518
 1802 AATGCTGAAAGGAGGAAAGGAAAGGAAATTCCTGCTTCTGATTTCTTATGATGAT 1861
 1519 AACCGGAGGAAAGGAGGAAAGGAAAGGAAATTCCTGCTTCTGATTTCTTATGATGAT 1578
 1862 ACAGTTGAGGTTTAAAGGAGGAAAGGAAAGGAAATTCCTGCTTCTGATTTCTTATGATGAT 1921
 1579 ACTGTAAGATCAAAAGGAGGAAAGGAAAGGAAATTCCTGCTTCTGATTTCTTATGATGAT 1638
 1922 AAATTTGCTGTTTATGATGATCAATCAAGCTTCAATGATGATGATGATGATGATGATGAT 1981
 1639 AAATTTGATGTTTGAAGAGGCTTATGAGGACCTTAAAGGAGGATTAATCTTGAAGAT 1698
 1982 TTGTTTGGAGCATTAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2041
 1699 CTTTGGAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1758
 2042 GGATGGGCAATGACGATGAGCATGTTATGAGCAAGAAAGCAAGTGAAGATCCAAT 2101

```

Db      1759 GGTGTTGGTAACGCTAGCAGCATGTCAAGAAACAAATGTCAGCTGATCAAT 1818
Oy      2102 AAGAACTTCAAGCGGATGA----- 2121
Db      1819 CAACGGAACAAACGAGAGAAACCTCAGACGAAAACTGAGAGAAACCCCT 1878
Oy      2122 -----AGAGCAGTAGAG 2134
Db      1879 CGAGAGAGAAACCCGAAAGCAGAAACAGAGTCTCCAAACCAACAGAGAACCAAGA 1938
Oy      2135 GAAACACCTGCTGAGCCAGAACTCCCTCAAGTAGAGACTGAAAAAGTAGAACCCCACTC 2194
Db      1939 GAATCACCAGAGAGATCAGAAAGAACCTCAGCTGACAGCTGAAAGAGTGAAGAAAACTG 1998
Oy      2195 AAAGAGCAGAAAGTTTGTCTTCCGAAAGTAAGAGATTTAGTCTGAAAGCCATGCAACA 2254
Db      1999 AGAGAGGCTGAAGATTTACTTGGAATAATCCAGATCCAAATTATCAAGTCCAAATCCAAA 2058
Oy      2255 GAAACTTAGCTGTTTGAATAATTTGACTCTTCAATATGATACAAATAGTATTC 2314
Db      2059 GAGACTTTCACAGATTAAAAATATTTACTATTGGCACCAGACCAACATCTATT 2118
Oy      2315 ATGCGAGAGCAGAAAAATTACTTCTGCTTTGTTAAAGAGATTA 2358
Db      2119 ATGCGAGAGCTGAATAAACTATTGCTTTATTAAGAGAGATTA 2162

```

RESULT 22

AAV52376
ID AAV52376 standard; DNA; 2359 BP.

AC AAV52376;

DT 23-OCT-1998 (first entry)

DE Streptococcus pneumoniae genome fragment SEQ ID NO:243.

XX Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;

KW computer readable medium; vaccine; pharmaceutical composition; ds.

XX Streptococcus pneumoniae.

PN WO9818931-A2.

XX 07-MAY-1998.

PF 30-OCT-1997; 97WO-US19588.

PR 31-OCT-1996; 96US-0029960.

XX (HUMA-) HUMAN GENOME SCI INC.

PI Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,

PI Kunsch CA, Rosen CA;

DR WPI, 1998-272225/24.

XX Computer-readable medium with recorded Streptococcus pneumoniae

PT polynucleotide sequences - useful in diagnostic kits and assays, and

PT pharmaceutical compositions and vaccines for Streptococcus

PT pneumoniae

PS Claim 1; Page 1265-1266; 1409pp; English.

XX The present invention describes a computer readable medium which has

CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)

CC recorded on it, or a representative fragment or a sequence at least 95%

CC identical to SEQ ID NO:1 to 391. The nucleotide sequences depicted in

CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from

CC Streptococcus pneumoniae. The present invention also describes an

CC isolated nucleic acid molecule encoding a homologue of any of the

CC fragments of the S. pneumoniae genome (SEQ ID NO:1 to 391) where the

CC nucleic acid molecule is produced by a process comprising: (a) screening
CC a genomic DNA library using as a probe a target sequence defined by any
CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC library which contain sequences that hybridize to the target sequence and
CC isolating the nucleic acid molecules from the members; or (b) isolating
CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC molecules whose nucleotide sequence is homologous to amplification
CC primers derived from the fragment of the S. pneumoniae genome to prime
CC the amplification and isolating the amplified sequences. The computer
CC readable medium can be used in a computer-based system for identifying
CC fragments of the S. pneumoniae genome of commercial importance, or
CC expression modulating fragments of the S. pneumoniae genome. Products
CC from the present invention can be used in diagnosis kits and assays, and
CC pharmaceutical compositions and vaccines for S. pneumoniae.

XX Sequence 2359 BP; 786 A; 451 C; 511 G; 611 T; 0 other;

Query Match 27.4%; Score 653.8; DB 19; Length 2359;

Best Local Similarity 68.3%; Pred. No. 1.1e-154;

Matches 1005; Conservative 0; Mismatches 413; Indels 54; Gaps 5;

```

Oy      1 TTCTTAGAGTTGGAGCTGATCAAGCTAGAACGCTTAAGGAAA---TAATCGTTTC 57
Db      939 TTCTTAGAGCTTGAGAGCTTAACAAGCTGTCAGAGATTAAGAAAGCTTAATCGAGTTGC 998
Oy      58 CTATATAGATGAAGAAACGAGCGACCAAAAAACGAGAAATTGTACTCTGATGAGTTAG 117
Db      999 TTATATAGATGATGATCAGCTGTCGTCAAAGCGAAAACTTGACACCAAGATGAAGTCAG 1058
Oy      118 CAAGCTGAAGAAATCAATGCTGAGCAATGTCATCAAGATTAACAGACCAAGCTATGT 177
Db      1059 TAAAGAGGAGGAGATCAACGCCGAAACAAATTGTATCAAGATTCAGATCAAGGTTATGT 1118
Oy      178 CACTTCACATGCGCAGCACTATCATTTCAATGTAAGTGAAGTTCCTATGACGCTATCAT 237
Db      1119 GACCTTCATGAGAACCATATCATTTCAATGTAAGGCAAGGTTCTTATGATGTCATCAT 1178
Oy      238 CAGTGAAGAAATTTACTCATGAAAGATCCAACTATTAAGCTTAAGAGATATTTGTTAA 297
Db      1179 CAGTGAAGAGCTCTCATGAAAGATCCGAATTTACATGTAAGATTCAGACATTTGTCAA 1238
Oy      298 TGAGTCAAGGCTGATATGTTATCAAGTATGATGAAATAATCATGATTTTACTTTAAGA 357
Db      1239 TGAATCAAGGCTGATATGTTATCAATTAAGTAAAGGTTAAATCATGATTTTACTTTAAGA 1298
Oy      358 TGCTGCCACGCGGATTAACGCTGACAAAAGAGAAATCAATGACAAAACAAAGACA 417
Db      1299 TGCAGCTCATCGGATTAATTTGAGACAAAAGAGATTAAGCTCAAGACAGAGAACG 1358
Oy      418 TAGTCAACATGCTGAAGGTGGAACCTCCAGAAACGATGCTGCTTGGCAGCGTTTC 477
Db      1359 CAGTCATATATATACT-----CAAGACAGATTAATGCTTGTGTCGACGACAGAGC 1409
Oy      478 GCAAGAGCGTATATCAAGATGATGTTATATCTTTATATCTTCTGATATCATAGAGA 537
Db      1410 CCAAGAGCGTATATCAAGATGATGTTATATCTTCAATGATCATGATATCATTTAGAGA 1469
Oy      538 TACTGATGATGCTATATGCTTCTCATGAGATCATTAACATTAACCTTCTTAAGATGA 597
Db      1470 CACGGGATGCTATATATGCTTCTCATGAGGACCATTAACATTAACCTTCTTAAGATGA 1529
Oy      598 GTTATCAGCTGAGAGTTGGCTGTCGAGAAAGCTTCTCATCTGCTGAGAGAAATCTGTC 657
Db      1530 GTTATCAGCTGAGAGTTAGCTGTCGAGAAAGCTTATTTGGAATG----- 1574
Oy      658 AATTTCAAGAACTTATCGCCGACAAAATAGCATTAACCTTCAAGAACAAATGCGGATAC 717
Db      1575 -GAAGCAGGAGATCTGTCCTTCTTCAAGTTCTATTAATGAACAAATCCAGCTCAACCAA 1633
Oy      718 TTCTGTAGCAATCCAGAACTCAAAATATCTAACAACAAGCAACAACGACATTAACAG 777
Db      1634 GATTGTAGAGAACCAACATCTAGCTGT-----CACTCAACTTA 1673

```



```

Db      2368 CATATGACCCATAGCCATGATTAATAAAGATAGTTTGTCTGAAGCTGAGAGCGGCA 2427
QY      1706 GCTCAAGCCTACTATAAGAAAAGGTAATCCATCCATCCAGAGCGAGATGTTAAA 1765
Db      2428 GCCAGGCTTATGCTTAAGAGAAAGGTTGACCCCTCTCTCGACAGACCATCAGATTCA 2487
QY      1766 GCMAATCCAACTGAGATAGTGACAGCTATTTTCAATCGTGTGAAAAGGGGAAAAAGCA 1825
Db      2488 GGAATATCTGAGGCAAAAAGACAGAGCTATCTCAACCGGTGAAAAGCGCTAAGAG 2547
QY      1826 ATTCACCTGCTTCCAGCTTCCATATATGTTGAGCAATACAGTTGAGTTAAAAAGGTAT 1885
Db      2548 GTGCCACTTGATGCTATGCTTACATCTTCAATATACGTAGAAATCAAAAAAGCGTAGT 2607
QY      1886 TTGATTATTCCTCATAGGATCATATACATATATTAATTTGCTGGTTGATGATCAC 1945
Db      2608 TTAACTATACCTCATATTAAGACCATTAACATCAAAATTTGAGTGGTTGACGAGGC 2667
QY      1946 ACATCAAAAGCTCCAAATGGCTATACCTTGAAGATTTTGTTCGACAGATTAAGTACTAC 2005
Db      2668 CTTTATGAGGACCTTAAGGGATATCTTTGAGATCTTTTGGCGACTGTCAAGTACTAT 2727
QY      2006 GTAGAACACCTGACGAAGCTCCACATTTCTAATGATGAGTGGGCAATGCCAGTGCAT 2065
Db      2728 GTCCAAACATCCAAAGCAAGCTCCGATTCAGATTAATGTTTGTGAACGCTAGCACCAT 2787
QY      2066 GTGTGAGCAAGAAAGACCAAGTGAAGATCCAAATTAAGAACTTCAAAAGCGATGAAG 2125
Db      2788 GTTCAAGAAACAAAATATGCTACACTGATACCAATCAAAACGAAACCAAGCGAGAG 2847
QY      2126 ----- 2125
Db      2848 AAACCTCAGACGAAAACTGAGAGAAAGAACCCCTCGAGAAAGAAACCAAAAGCGAG 2907
QY      2126 -----CCAGTAGAGGAAACCTCTGAGCGAGAA 2155
Db      2908 AAACGAGAGTCTCAAAAACCAACAGAGAGAACGAAAGAAATCACAGAGAAATCAGAA 2967
QY      2156 GTCCCTCAAGTAGACAGTGAAGAAAGTGAAGCCCAACTCAAAAGACGAGATTTTGGTT 2215
Db      2968 GAACCTCAGCTGACAGTGAAGAAAGGTTGAAGAAAGTGAAGAGGCTGAAGATTTACTT 3027
QY      2216 GCGAAAGTAGAGATTTAGTCTGAAAGCCAAATGCAAGAAACTAGTCTGTTTACGA 2275
Db      3028 GGAAGAAATCCAGATCCAAATTAATCAAGTCCAAATGCAAGAGACTCTCAGAGGATTA 3087
QY      2276 AATTAATTTGACTCTTCAATTAATGATGAACAATAGTATCAGTGCAGAGCAAGAAATTA 2335
Db      3088 AATTAATTTACTATTTGGCACCCAGCAACAATATCTATTATGGCAGAGCTGAAAACTA 3147
QY      2336 CTGGCTGTTAAAGAAAGTAA 2358
Db      3148 TTGGCTTATTAAGAGAGTAA 3170

RESULT 24
AAV27414
ID      AAV27414 standard; DNA; 1342 BP.
XX      AAV27414;
XX      02-OCT-1998 (first entry)
XX      Streptococcus pneumoniae SPI03 nucleotide.
XX      Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
XX      detection; pneumonia; otitis media; meningitis; ss.
XX      Streptococcus pneumoniae.
XX      Key Location/Qualifiers
XX      CDS 2..1342
XX      FT /*tag= a

```

```

FT      /product= "SPI03"
FT      /note= "no stop codon given"
XX      MO9818930-A2.
XX      07-MAY-1998.
XX      30-OCT-1997; 97WO-US19422.
XX      31-OCT-1996; 96US-0029960.
XX      (HUMA-) HUMAN GENOME SCI INC.
XX      ChOI GH, Hromockyj A, Johnson LS, Kunsch CA;
XX      WPI; 1998-272224/24.
XX      P-PSDB; AA61228.
XX      Nucleic acid encoding antigenic peptide(s) from Streptococcus
XX      pneumoniae - or their epitope-containing fragments, useful in
XX      protective or therapeutic vaccines, and for diagnosis
XX      Claim 1; Page 85; 118pp; English.
XX      The present sequence encodes a protein from Streptococcus pneumoniae.
XX      The nucleic acid sequence encoding the Streptococcus pneumoniae protein
XX      can be useful in vaccines for inducing protective antibodies against
XX      Streptococcus pneumoniae, for treatment or prevention of infection e.g.
XX      pneumonia, otitis media or meningitis. Probes based on the nucleic acid
XX      are used to detect Streptococcus infection (by usual hybridization or
XX      amplification methods), also for isolating Streptococcus genes or their
XX      allelic variants. The protein can be used similarly to detect specific
XX      antibodies in standard immunoassays, especially for diagnosing or
XX      monitoring infections. Antibodies which bind the protein are used to
XX      detect corresponding antigens, to purify the protein and for passive
XX      immunisation (optionally coupled to a toxin). Vaccines are administered,
XX      e.g. by injection, orally or through the skin, typically at 0.01-1000
XX      (especially 10-300) mu g/ml per dose.
XX      Sequence 1342 BP; 451 A; 270 C; 261 G; 360 T; 0 other;
XX      Query Match 16.1%; Score 385.4; DB 19; Length 1342;
XX      Best Local Similarity 65.3%; Pred. No. 5.3e-87;
XX      Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;
QY      37 TAAAGAAATTAATGTGTTTCTTATATAGTGAAGAAACGAGCGGAAAAACGGAGAA 96
Db      28 TAAAGACAAATTAATGTGTTTCTTATATGATGCGAGCTCAAGTCAAGAAAGTGA 87
QY      97 TTGACTCTGATAGGTTAGCAAGCGTGAAGAAATCAATGCTGAGCAAAATCGTATCAA 156
Db      88 CTTGACACACAGACAGGTTAGCCGAAAGAAAGAAATTCAGGCTGACAAATTTGTAATCA 147
QY      157 GATTAACAGACCAAGGCTATGCTCACTTCAATGCGACCACTATCATTAATTAAGTAA 216
Db      148 AATTACAGATCAGGCGTATGTAACGTACACAGGACCACTATCATTAATTAAGTAA 207
QY      217 GGTTCCTTAGAGCGCTATCACTAGTGAAGAAATTAATGTAAGAAATCCAAATTAAGCT 276
Db      208 AGTTCCTTATGATGCTCTTCTTATGTAAGAACTCTTATATGAAGATCAAACTATCACT 267
QY      277 AAAAGATGAGGATTTGTTAATGAGTCAAGGGGATGATGTTTCAAGGATGAGTGA 336
Db      268 TAAAGAGCGCTGATATTTCAATGAAGTCAAGGGGTTATTAATCAAGGATGAGTGA 327
QY      337 ATACTATGTTTACCTTAAGATGCTGCCACGCGGATTAACGTCCGTACAAAAGAGAAAT 396
Db      328 AATATTATGCTACTCTGAAGATGACGCTACATGCTGATTAATGTTGAAGTAAAGTGAAT 387
QY      397 CAATGACAAAACAAAGACATAGTCAACATCGTGAAGGTGGAAGTCAAGAAAGATGCG 456
Db      388 CAATCGTCAAAAACAAAGACATGTCAAAGATTAATGAGAAAGTGA-----ACTC 435

```

QY 457 TCGTGTGCTTGGCAGCTTTCGCAAGCGCTAATACAGATGATGTTATATCTTTAA 516
 DB 436 TAATGTGCTGTAGCAAGGCTCTCAGGAGCATATACGAATAATGATGTTATGCTTTAA 495
 QY 517 TCGTGTGATATCATAGAGATATCTGATGTTATGCTTATGCTTCCATGAGATATTA 576
 DB 496 TCCAGTGTATATATGAAAGATACGGGTATGTTATGCTTATGCTTCCATGAGGCTACAA 555
 QY 577 CCATTACATTCCTAAGATGAGTTATCAGTACGAGTGGCTGCTGCAAGACCTTCT 636
 DB 556 TCACATACATTCCTCAAAAGGATTTATCTGTAGTGAATTAAGCAGCGCTAAAGCATCT 615
 QY 637 ATCTGTGTCAGGAATCTGTCAAAATTCAGAACCTATGCGCCGCAAAATAGCGATTAAC 696
 DB 616 GCGTGAAGAAAATATGCAACCGAGTCAAGTTAAGCTATCTTCAACGCTAGTGACAA 672
 QY 697 TTCAGAACAAACTGGGTACTTCTGTAGCAATCCAGAACATCAAAATCTTAACACAG 756
 DB 673 -----TMAACGCCAATCTGT 687
 QY 757 CAACAACAGCACTAATACAGTCAAGCAGTCAAAAGTAAATGACATGATAGTCTCTGAA 816
 DB 688 AGCAAAAGATCACTACAGCAGCAGCAATAATCTGAAATCTCAGAGTCTTTTGA 747
 QY 817 ACAGCTTCAAACTGCTTGTAGTCAACAGACATGATGATGATGCTTGTCTTTGA 876
 DB 748 GGAACCTATATGATCCTAGGCGCCCAAGTTACAGTGAATCAGATGCGCTGTCTTTGA 807
 QY 877 TCCAGCACAATATCAAGTGCAGACGCTAGAGTGTGACAGTCCACAGAGATCATTA 936
 DB 808 CCTCTAGATATATCAAGTGCAGACCAATGAGATGATGATGCGCATGCGGACCATTA 867
 QY 937 CCAGTTCATCCCTTACTCTCAATATGCTGAATTTGAAGAAGCAATGCGTATTAATTC 996
 DB 868 CCAGTTCATCCCTTACTCTCAAGCAGCTTCTGCTTAGAAGAAAGATGCGAATGCTG 927
 QY 997 CCTTGT 1003
 DB 928 TATCAGT 934
 RESULT 25
 ABQ84882
 ID ABQ84882 standard; DNA; 1342 BP.
 AC ABQ84882;
 XX
 DT 04-SEP-2002 (first entry)
 XX
 DE S. pneumoniae SPI03 nucleotide sequence SEQ ID NO:181.
 XX
 KW Streptococcus pneumoniae; epitope; vaccine; antigenic protein;
 KW antibacterial; Streptococcal infection; detection; gene; ds.
 OS Streptococcus pneumoniae.
 XX
 PN US2002061545-A1.
 PD 23-MAY-2002.
 XX
 PF 22-JAN-2001; 2001US-0765272.
 XX
 PR 30-OCT-1997; 97US-0961083.
 PA (CHOI/) CHOI G H.
 PA (KUNS/) KUNSCH C A.
 PA (BARA/) BARASH S C.
 PA (DILL/) DILLON P J.
 PA (DOUG/) DOUGHERTY B.
 PA (FANN/) FANNON M R.
 PA (ROSE/) ROSEN C A.
 PI Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;

PI Rosen CA;
 XX
 DR MPI: 2002-479261/51.
 DR P-PSDB; ABP5467.
 XX
 PT New Streptococcus pneumoniae antigens, useful for detecting
 PT Streptococcus and for preventing or attenuating disease caused by
 PT Streptococcus infection -
 PS Claim 1; Page 45; 70pp; English.
 XX
 CC ABQ84792 to ABQ84904 represents nucleic acids which encode the
 CC Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.
 CC The S. pneumoniae antigens have antibacterial activity and can be
 CC used in vaccines. The S. pneumoniae antigens can also be used to
 CC prevent or attenuate a Streptococcal infection in an animal. The
 CC polynucleotides encoding the S. pneumoniae antigens can be used to
 CC detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent
 CC primers used in the cloning of S. pneumoniae ORFs (open reading frames)
 CC which are used in an example from the present invention.
 XX
 SQ Sequence 1342 BP; 451 A; 270 C; 261 G; 360 T; 0 other;
 Query Match 16.1%; Score 385.4; DB 24; Length 1342;
 Best Local Similarity 65.3%; Pred. No. 5.3e-87;
 Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;
 QY 37 TAAAGAAAATATCTGTGTTCTTATATGATGAAACAGAGACCAAAACGAGAA 96
 DB 28 TAAAGCAATATATCTGTCTCTTATGTGATGAGCGCAAGTCAAGTCAAGAAAGTGA 87
 QY 97 TTTGACTCCTGATGAGTTAGCAAGCGTGAAGGAATCATGCTGCAAAATGCTATCA 156
 DB 88 CTGACACCAAGCCAGTTAGCCAGAAAGAAATTAAGGCTGAGCAATGTATCA 147
 QY 157 GATTAACAGACCAAGGCTATGTCATCTTCAATGCGCAGCACTATCATTTTCAATGTA 216
 DB 148 AATTACAGATCAGGCTATGTAAGTCAACAGGTAACCATATCATTTATTAAGGGA 207
 QY 217 GGTTCCTTATGACGCTATCATGAGTAAGGAATTAATCAATGAAGTCCAAATTAAGT 276
 DB 208 AGTTCCTTATGATGCGCTTCTTATGAAAGAACTCTTGAATGAAGATCCAAATCATCACT 267
 QY 277 AAAAGATGAGTATGTTTAAATGAGTCAAGGCTGATATGTTATCAAGTATGAGAA 336
 DB 268 TAAAGACGCTGATATGTCATGAAGTAAGGCTGATATCATCAAGTGAAGAA 327
 QY 337 ATACTATGTTTACCTTAAGATGCTGCCACGCGATTAACGCTGTAACAAAGAGAAAT 396
 DB 328 AATATATGCTTACCTGAAGATGAGCTCATGCTGATATATGTTCAACTAAAGATGAAT 387
 QY 397 CAATGACAAAACAGAGCATATGCAACATCTGTAAGGTGAAGTCCAAAGAAAGATG 456
 DB 388 CAATGTCAAAACAGAGCATATGCAAGATATATGAGAGTTA-----ACTC 435
 QY 457 TCGTGTGCTTGGACGCTTGGCAAGAGCGTATCTACAGATGATGTTATATCTTTAA 516
 DB 436 TAATGTGCTGTAGAGGCTCTCAGGAGCATATACGAATAATGATGTTATGTTTAA 495
 QY 517 TCGTGTGATATCATAGAGATATCTGATGTTATGCTTATGCTTCCATGAGATATTA 576
 DB 496 TCCAGTGTATATATGAAAGATACGGGTATGTTATGCTTATGCTTCCATGAGGCTACAA 555
 QY 577 CCATTACATTCCTAAGATGAGTTATCAGTACGAGTGGCTGCTGCAAGACCTTCT 636
 DB 556 TCACATACATTCCTCAAAAGGATTTATCTGTAGTGAATTAAGCAGCGCTAAAGCATCT 615
 QY 637 ATCTGTGTCAGGAATCTGTCAAAATTCAGAACCTATGCGCCGCAAAATAGCGATTAAC 696
 DB 616 GCGTGAAGAAAATATGCAACCGAGTCAAGTTAAGCTATCTTCAACGCTAGTGACAA 672
 QY 697 TTCAGAACAAACTGGGTACTTCTGTAGCAATCCAGAACATCAAAATCTTAACACAG 756

ID		AAAA7603 standard; DNA; 1455 BP.
XX AC	AAA47603;	
XX DT	20-OCT-2000	(first entry)
XX XX		
DE		Recombinant variant of Sp36 gene (Sp36E) of S. pneumoniae.
KW KW		Streptococcus pneumoniae; infection; vaccine; coiled coil region; histidine triad residue; Sp36; antibody; otitis media; nasopharyngeal infection; bronchial infection; bronchitis; sepsis; meningitis; lobar pneumonia; ds.
XX OS		Streptococcus pneumoniae.
XX FH	Key	Location/Qualifiers
FH FT CDS	1..1455 /*tag= a /product= Sp36f polypeptide	
NN NN	WO200037105-A2.	
PD PD	29-JUN-2000.	
XX PF	21-DEC-1999;	99WO-US30390.
XX PR	21-DEC-1998;	98US-0113048.
XX PA	(MEDI-) MEDIMUNE INC.	
PI PI	Johnson LS, Koenig S, Adamou JE;	
XX DR	WPI; 2000-452129/39.	
DR DR	P-PADB; AAB01467.	
PT PT	Vaccine useful for prophylaxis and treatment of pneumococcal infections such as otitis media, nasopharyngeal and bronchial infections. Comprises Streptococcus pneumoniae proteins	
XX PS	Disclosure; Page 60-61; 70pp; English.	
CC CC	Although a number of proteins have been suggested as being involved in the pathogenicity of Streptococcus pneumoniae, there still remains a need to identify polypeptides having epitopes in common from various strains of S. pneumoniae in order to utilize such polypeptides in vaccines to protect against a wide variety of S. pneumoniae. New vaccine compositions are described which comprise a Streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino acids in length that comprise at least one histidine triad residue (HXxHKx) or a coiled-coil region, or an antibody directed against these features. The vaccine is useful in protecting against infection by Streptococcus pneumoniae. The vaccine composition comprising antibodies to its useful for passive immunization for treating Pneumococcal infections which includes otitis media, nasopharyngeal and bronchial infections.	
SQ SQ	Sequence 1455 BP; 491 A; 287 C; 285 G; 392 T; 0 other;	
OY Db	Query Match Best Local Similarity Matches 631; Conservative	16.1%; Score 385.4; DB 21; Length 1455; Pred.No.5.5e-87; Mismatches 276; Indels 60; Gaps 2
Oy Db	37 TAAGAAATTAATCGTGTTCCTAATAATGATGAACCAAGCAGCACCAAAGGAGAGA 96 99 TAGGACAATATCTGTCTCTTTATTGTGATGGCAGGCCAGTGCAAGAACAAAGTAGAAA 158 97 TTGAATCTCTGATGAGGTGTAGCAAGCGTGAAGAAATCAATGCTGAGCAAAATGTCATCAA 156 159 CTTGACACACAGACCAGGTGTAGCCAAGAAAGMAAGAAATTCAGCTGAGCAAAATGTGAATCAA 218 157 GATAACACACCAAGGCTATGTACTTCCCATATGCATGCGACCACTATTCATTATTAAGAGSTA 216 219 AAATTCAGATCAGGGCTATGTATAGGTCAACAGGTGACCACTATTCATTACTATTAATGGAGA 278	

Qy	217	GGTCCCTTATGACGGATCATCATAGGAAGATTAAGTCAATGAAGATCCAAACTTAAGCT	276
Db	279	AGTTCTTATGATGCCCTCTTTAGAGAAAGAACTCTTATGAGAGATCCAACTTCACT	338
Qy	277	AAAAGATGAGATATTGTTTAATGAGTCAAGGGTGAATATGTTATCAAGTAGATGAAA	336
Db	339	TAAAGACGGTATATGTCAATGAAGTCAAGGGTGGTATATCATCAAGGTCGATGAAA	396
Qy	337	ATATCTATGTTTACCTTAAGAGATGCTGCCACGGGATTAAGTCCGTCACAAAAGGAAAT	396
Db	399	ATATTTATGCTACCTGAAGAGTGCAGCTCATGCTGATATATTTGAACTAAAGATGAAAT	458
Qy	397	CAATGCACAAAACCAAGACATAGTCACATCGGAAGGTGAAGCTCCAGAAACGATGG	456
Db	459	CAATCTGCAAAAACCAAGACATGTCAAAAGATTAAGAGAGTTA-----ACTC	506
Qy	457	TGCTGTGCTTGGCAGCTTGCAGAGAGCGCTATATCAAGATAGTGTATATCTTTAA	514
Db	507	TAAATGTTGGTATGCAAGGTCCTCAAGGACGATATATCGACAATAGTGGTATATGCTTTAA	566
Qy	517	TGCTTCTGATATCATAGAGATACCTGGATGCTTATATGCTTCTCATGAGATCATTA	576
Db	567	TCCAGCTGATATTTATCGAAGATACGGGTAAAGCTTATATCTTCTCATGAGAGTCACTA	626
Qy	577	CCATTCATTTCTTAAGATGAGTATATCATGCTAGAGAGTGGCTGTGAGAGGCTTCT	636
Db	627	TCACTTCATTTCCCAAAAGCGATTTATCTGCTATGTAATAGACAGACGTAAAGACATCT	686
Qy	637	ATCTGTGTCAGAGAAATCTGTCAAAATTCAGAACTTATCGCCGACAAATAGCGATAAC	696
Db	687	GGCTGGAAAAAATATGCAACCGAGTCAGTTAAGCTATCTTCAACAGCTAGTGACAA---	743
Qy	697	TTCAAGAACAACTGGGTACTTCTGTAGCAATCCAGAACTACAAATCTAACACAG	756
Db	744	-----TAAACGCAATCTGT	758
Qy	757	CAACAACAGCAACACTTAACAGTCAAGCAAGTCAAAAGTAAATGACATTGATAGTCTCTGAA	816
Db	759	AGCAAAAGATCAACTAGCAACCGACGAAATTAATCTGAATAATCTCCAGAGCTCTTTGAA	818
Qy	817	ACAGCTCTCAAACTGCTTGAATGCAACGACATGATGAATCTGATGCTTGTCTTTGA	876
Db	819	GGAACCTCTATGATTCACCTAGCGCCCAACGTTACAGTAATCAAGATGGCTGGCTTTGA	878
Qy	877	TCAGACAAATATCAAGTCCAAAGCTAGAGGTGTTCAGATGCCACAGGAGATCATTA	936
Db	879	CCCTGTAAAGATTATCAAGTCGACACCAATAGGATGGATTCGGATTCGGATGCCGACATTA	938
Qy	937	CCACTTCATCCCTTACTCTCAAAATGCTGAATTTGGAAGAAAGAAATGCTGATATTATTC	996
Db	939	CCACTTATTTCTTACAGCAAGCTTCTGCTTAGAAGAAAAGATTCGCAAGATGTGCC	998
Qy	997	CCTTGCT 1003	
Db	999	TATCAGT 1005	

RESULT 28

AAA05473 standard; DNA; 1455 BP.

AAA05473:

AAA05473:

24-MAY-2000 (first entry)

Streptococcus pneumoniae type 4 nucleotide sequence 4138.1.

Streptococcus pneumoniae; vaccine; screening; protein antigen;

antibacterial; antiinflammatory; meningitis; infection; diagnosis;

pneumococcal disease; ds.

Streptococcus pneumoniae.

XX WO200006737-A2.
 XX 10-FEB-2000.
 XX 27-JUL-1999; 99WO-GB02451.
 XX 27-JUL-1998; 98GB-0016337.
 XX 19-MAR-1999; 99US-0125164.
 XX (MICR-) MICROBIAL TECHNIQS LTD.
 XX Gilbert CEG, Hansbro PM;
 XX WPI; 2000-195300/17.
 XX
 PT New Streptococcal protein, useful as a vaccine, for diagnosis of
 PT pneumococcal diseases and for screening agents capable of antagonizing
 PT or inhibiting expression of the protein
 XX
 PS Claim 7; Page 41-42; 108pp; English.
 XX
 CC AAY81501 to AAY81679 represent specifically claimed protein sequences
 CC isolated from Streptococcus pneumoniae. AA05407 to AA05590 represent
 CC specifically claimed nucleotide sequences isolated from S. pneumoniae.
 CC The sequences have antibacterial and anti-inflammatory properties.
 CC The protein sequences, and fragments of them, are useful as immunogens
 CC and/or antigens. The nucleotide sequences can be used in vaccines and in
 CC diagnostic assays. The proteins and nucleotides can be useful for the
 CC detection and diagnosis of S. pneumoniae. The protein sequences are also
 CC useful for screening an agent capable of antagonizing, inhibiting or
 CC interfering with the function or expression of the proteins in which the
 CC agent is useful for treatment or prophylaxis of S. pneumoniae infection
 CC and meningitis. AA05591 to AA05614 represent primers used in the
 CC exemplification of the present invention.
 CC
 XX
 XX Sequence 1455 BP; 491 A; 287 C; 285 G; 392 T; 0 other;

Query Match 16.1%; Score 385.4; DB 21; Length 1455;
 Best Local Similarity 65.3%; Pred. No. 5.5e-87;
 Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

QY 37 TAAGAAAATAATCGTGTTCCTATATAGTGAAGAAACGAGCAAGCAAAACGAGAA 96
 DB 99 TAAAGACATATATGCTGTCTTATGTGATGGACGACGATCAAGTGAAGAAAGGABAA 158
 QY 97 TTGACTCTGATGAGTTAGCAAGCGTGAAAGAAATCATGTGTGACAAATGTCATCA 156
 DB 159 CTGGACACGACGAGGTTAGCCAGAAAGAAAGAAATTCAGGCTGAGCAAAATGTTATCA 218
 QY 157 GATAACAGACCAAGGCTATGTCACCTCAATGCGGACCACTATCTTTTCAATGTGTA 216
 DB 219 AATTACAGATCAGGCGTATGTACGTCACGCGTACCACTATCTTCTATATGGGA 278
 QY 217 GGTTCCTATGACGCTATCATGATGAGAAATTAATCTGAAGATCCAACTAATAGCT 276
 DB 279 AGTTCCTATATATGCGCTCTTTAGTGAAGAACTCTGATGAAGATCCAACTATCACT 338
 QY 277 AAAAGATGAGATATTTTAATAGGTCAAGGGTGATGTATGTTCAAGATGATGAA 336
 DB 339 TAAAGACGCTGATATGTCATGAAGTCAAGGGTGTATATCATCAAGTGTGATGAA 398
 QY 337 ATACATGTTTACCTTAAGATGCTGCCACCGGATTAAGTCCGTACAAAGAGAAAT 396
 DB 399 ATATTATGTTACCTGAAGATGACGCTCATGCTGATATATGTTCAAACTAAGATGAAT 458
 QY 397 CAATCGCAAAAACAGACATAGTCAACATCGTGAAGGTGAATCCAAAGAAAGATG 456
 DB 459 CAATGCTAAAAACAGAACATGTCAAAGATATAGAAAGGTA-----ACTC 506
 QY 457 TGTGTTGCTTGGACGCTTGCAGAGACGCTATATCAAGATGATGTTATATCTTTAA 516
 DB 507 TAAATGCTGTAGCAAGGTCTCAGGAGACGATATACCAATGATGTTATGCTTTAA 566

QY 517 TGTCTGATATCATAGAGATACCTGATGATGCTTATATGTTCCATGAGATCATTA 576
 DB 567 TCCAGCTGATATATGCAAGATACGGGTAAATGCTTATATGTTCTCATGAGATCACTA 626
 QY 577 CCATTACATTCCTAAGATGATGATATCAAGTACGAGAGTGGCTGCTCAGAAAGCTTCT 636
 DB 627 TCATCATATTTCCCAAAAGCATTTATCTGTAGTGAATTAGCAGACCTAAAGACATCT 686
 QY 637 ATCTGTGAGAAATCTGCAATTCGAATTCGAACCTATCGCGCAAAATAGCGATTAAC 696
 DB 687 GGTGGAATAATATGCAACCGAGTCACTTAAGCTTATCTTCAACAGACTAGTGA--- 743
 QY 697 TTCAAGAACAAACTGGTACCTTCTGTAGCAATCCAGAACTAACAATATCAACAG 756
 DB 744 -----TAAACGCAATCTGT 758
 QY 757 CAACAACAGCAACTACAGTCAAGCTCAAGCAAGTAAATGACATTTATGCTCTTGA 816
 DB 759 AGCAAAAGATCAACTAGCAAGCCAGCAAAATTAATCTGAAATCTCCAGAGCTTTGAA 818
 QY 817 ACAGCTTACAAACTGCTTGAAGTCAAGCAAGTGAATCTGATGCGCTTGTCTTGA 876
 DB 819 GGAAGCTATGATTCACCTAGCGCCCAAGCTTACAGTGAATCAGATGAGCTGCTTTGA 878
 QY 877 TCCAGCAAAATCACAAGTGAAGAGTGTGCAAGTCCAGCAAGAGATCATTA 936
 DB 879 CCCGCTAAGATATATCATGCTGATACCAAAATGAGATTGCCATGCCAGCACTTA 938
 QY 937 CCATTCATCCCTTACTCTCAAAATGTGTAATGGAAGAAAGAAATCCCTGATTAATCC 996
 DB 939 CCATTTATTTCTTACAGCAAGCTTCTGCTTGAAGAAAGATGCCAAGATGTGCC 998
 QY 997 CCTTGT 1003
 DB 999 TATCAGT 1005

RESULT 29

AA291804
 ID AA291804 standard; DNA, 1455 BP.

XX AA291804;
 DT 02-JUN-2000 (first entry)
 XX
 DE Streptococcus pneumoniae DNA sequence ID128.
 XX
 KW Streptococcus pneumoniae infection; immunogen; antigen; diagnosis; AIDS;
 KW bacterial pneumonia; aplenia; heart disease; lung disease; alcoholism;
 KW kidney disease; diabetes; immunosuppressive disorder; otitis media;
 KW pneumococcal septicemia; sinusitis; meningitis; therapy; ss.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200006738-A2.
 PD 10-FEB-2000.
 XX
 PF 27-JUL-1999; 99WO-GB02452.
 XX
 PR 27-JUL-1998; 98GB-0016336.
 PR 19-MAR-1999; 99US-0125329.
 XX
 PA (MICR-) MICROBIAL TECHNIQS LTD.
 XX
 PI Le Page RWF, Wells JM, Hanniffy SB, Hansbro PM;
 XX WPI; 2000-195300/17.
 DR P-PSDB; AAY81708.
 XX
 PT Streptococcal proteins and polynucleotides useful for diagnosis,
 PT treatment and prophylaxis of bacterial infections

XX Claim 1; Page 39; 76pp; English.
 PS This sequence encodes a Streptococcus pneumoniae protein of the
 XX invention. The protein (or their homologues, derivatives and/or
 CC fragments) are useful as immunogens or antigens. Immunogenic or antigenic
 CC compositions comprising the proteins are useful as vaccines and also in
 CC diagnostic assays. The sequences are useful for the detection or
 CC diagnosis of S. pneumoniae infection, by contacting a sample to be tested
 CC with them. Agents capable of antagonising, inhibiting or interfering with
 CC the function or expression of the protein or polypeptide are useful in
 CC medical compositions in the treatment or prophylaxis of S. pneumoniae
 CC infection. As the sequences can be used to treat S. pneumoniae infection,
 CC they can be used to treat bacterial pneumonia, which has high rates in
 CC young children, the elderly, and in patients with predisposing conditions
 CC such as asplenia, heart, lung and kidney disease, diabetes, alcoholism,
 CC or with immunosuppressive disorders, especially AIDS. They can also be
 CC used to treat pneumococcal septicemia, otitis media, sinusitis, and
 CC meningitis.
 CC
 SQ Sequence 1455 BP; 491 A; 287 C; 265 G; 392 T; 0 other;

Query Match 16.1%; Score 385.4; DB 21; Length 1455;
 Best Local Similarity 65.3%; Pred. No. 5; 5e-87;
 Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

OY 37 TAAGAAAAATATCGTGTCTTCTATATAGTGAAGAAAAAGGAGGCAAAACGAGAA 96
 DB 99 TAGAGACATATATCGTGTCTTCTATATAGTGAAGAAAAAGGAGGCAAAACGAGAA 158
 OY 97 TTTGACTCTGATGAGTTAGCAAGCGTGAAGAAATCAATGCTGAGCAATGCTATCAA 156
 DB 159 CTGACACAGACAGGTTAGGCAAGAAAGAAATTCAGGCTGAGCAATGTTATCAA 218
 OY 157 GATAACAGACCAAGGCTATGCTTCAATGCGACCACTTATTTAATTAAGTGA 216
 DB 219 AATTACAGATCGGCTATGATGACGTCACGCTGACCACTTATTTAATTAAGTGA 278
 OY 217 GGTTCCTATGAGCGTATCATGAGTGAAGAAATTAATCAAGAAAGTATTAAGT 276
 DB 279 AGTTCCTATGAGCGTATCATGAGTGAAGAAATTAATCAAGAAAGTATTAAGT 338
 OY 277 AAAAGATGAGATATTTGTAATGAGTCAAGGTTGATATTTAATCAAGTATGAGAA 336
 DB 339 TAAAGACGCTGATATTTGTAATGAGTCAAGGTTGATATTTAATCAAGTATGAGAA 398
 OY 337 ATACTATGTTTACCTTAAGATGCTGCCACGCGATTAAGTCCGTAACAAAGAGAAAT 396
 DB 399 ATATTATGTTTACCTTAAGATGCTGCCACGCGATTAAGTCCGTAACAAAGAGAAAT 458
 OY 397 CAATGACAAAAAAGAGATGATGCAATGCTGAAGTGAACCTCCAAAGAGATG 456
 DB 459 CAATGACAAAAAAGAGATGATGCAATGCTGAAGTGAACCTCCAAAGAGATG 506
 OY 457 TCGTGTGCTTGGACGTTCCGAAAGAGATGATGCAATGCTGAAGTGAACCTCC 516
 DB 507 TAATTTGCTGAGCAAGGCTCAGGAGATGATGCAATGCTGAAGTGAACCTCC 566
 OY 517 TCGTGTGATATCATAGAGATGATGCTGAAGTGAACCTCCGTAACAAAGAGATG 576
 DB 567 TCCAGCTGATATTCAGAGATGATGCTGAAGTGAACCTCCGTAACAAAGAGATG 626
 OY 577 CCATTAACATCTCTAGAAATGATGCTGAAGTGAACCTCCGTAACAAAGAGATG 636
 DB 627 TCATTAACATCTCTAGAAATGATGCTGAAGTGAACCTCCGTAACAAAGAGATG 686
 OY 637 ATCTGTCGAGGAAATCTGTCAAAATTCAGAACTATCCGCAACAAATAGGATTAAC 696
 DB 687 GCGTGAAGAAATATGCAACCGAGTCAAGTATTTTCAACAGCTAGTGAACA--- 743
 OY 697 TTCAAGAACAACTGGTACTTCTGTGAAGCAATCCAGAACTACAAATATTAACAG 756
 DB 744 -----TAACACCAATCTGT 758

OY 757 CAACAAGACAACTAATAGTCAAGAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 816
 DB 759 AGCAAAAGATCAATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 818
 OY 817 ACAGCTTCAAACTGCTTGTAGTCAAGATGATGATGATGATGATGATGATGATGAT 876
 DB 819 GGAATCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 878
 OY 877 TCCAGCAAAATCAACAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 936
 DB 879 CCTGCTAAGATTAATCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 938
 OY 937 CCAGTTCATCTCTGCTCAATGCTGAATTTGAGAGAGAGATGCTGATTAATGCC 996
 DB 939 CCAGTTCATCTCTGCTCAATGCTGAATTTGAGAGAGAGATGCTGATTAATGCC 998
 OY 997 CCTTGT 1003
 DB 999 TATCAGT 1005

RESULT 30
 ABX06706
 ID ABX06706 standard; DNA; 3117 BP.
 XX
 AC ABX06706;
 XX
 DT 11-FEB-2003 (first entry)
 XX
 DE S. pneumoniae type 4 strain coding region #994.
 XX
 KW Genes; ds; bacterial meningitis; pneumonia; sepsis; otitis media;
 KW ear infection; anti-inflammatory; antibacterial; immunostimulant;
 KW auditory; respiratory; gene therapy; vaccine.
 OS Streptococcus pneumoniae type 4 strain.
 XX
 PN MO200277021-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 27-MAR-2002; 2002MO-IB02163.
 XX
 PR 27-MAR-2001; 2001GB-0007658.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Masignani V, Tettelin H, Fraser C;
 XX
 DR WPI; 2003-040579/03.
 DR P-PSDB; AB01419.
 XX
 PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
 PT or ear infection
 XX
 PS Claim 6; SEQ ID No 1987; 56pp; English.
 CC The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC ABS6454. Also included are an antibody which binds one of the
 CC proteins, treating a patient by administering the protein, DNA or
 CC antibody (in a composition), a kit comprising first and second primers,
 CC which are the nucleic acid cited above or fragments between nucleotides
 CC 8-100 of a sequence not defined in the specification, for amplifying a
 CC target sequence contained within a Streptococcus nucleic acid sequence,
 CC where the first primer is substantially complementary to the target


```

QY 37 TAAGAAAATAATCTGTTTCTTATATAGTGAACAAAGGACGCAAAAAACGAGAA 96
Db 99 TAAGACAATAATCTGTTCTTATGTGATGGACCCAGTCAAGTCAAGAAAGTGAAA 158
QY 97 TTGACTCCTGATGAGTTAGCAAGGTAAGGAATCAATGCTGAGCAAAATGTCATCA 156
Db 159 CTTGACACCAAGCCAGGTTAGCCAGAAAAGAAATTCAGGCTGAGCAAAATGTTATCA 218
QY 157 GATAACAGACCAAGGCTATGCTCATCTCAATGCGCAGCACTATCATTTATTAACATGTA 216
Db 219 AATTACAGATCAAGGCTATGTAACGTCACAGGTAACCTATCATTTATTAATGGA 278
QY 217 GGTCTCTATGACGCTATCATGATGAAGAAATTAATCAAGAAAGTCAAACTATAGCT 276
Db 279 AGTCTCTATGATGCCCCCTTATGTAAGAACTCTTATGAAGAAATCAAACTATCACT 338
QY 277 AAAAGATGAGATTTTGTATATGAGTCAAGGTTGATATGTTATTAAGTGAATGAAA 336
Db 339 TAAAGACGCTGATATTTGTCAATGAAGTCAAGGTTGTTATCATCAAGTCAATGAAA 398
QY 337 ATACTATGTTTACCTTAAGATGCTCCAGCGGTAACGTCCTGTAACAAAGAGAAAT 396
Db 399 ATATTATGCTACCTGAAAGATGCACTCATGCTATATGTTCAACTTAAAGATGAAT 458
QY 459 CAATGTCAAAAACAAGAACATGTCMAAGATATAGAAAGTTA-----ACTC 506
Db 457 TGCTGTGCTTGGCAAGTTCGCAAGAGACGCTATACATGATGATGTTATATCTTAA 516
QY 507 TAAATGTGCTGAGCAAGGCTCTCAGGACGATATACGAACAAATGATGTTATGCTTAA 566
Db 517 TGCTCTGATATCATAGAGATGCTGATGATGCTTATATGCTCCATGAGATCATTA 576
QY 567 TCCAGCTGATATATCGAAGATACGGATATGCTTATATGCTTCTCATGAGGTTACTA 626
Db 577 CCATTACATCTTAAGATGATGTTATCACTAGCAGATGCTGTCGCAAGACCTTCT 636
QY 627 TCACACTACATCCCAAAAGGATTTATCTGCTAGTGAATTAAGACGAGCTTAAAGCATCT 686
Db 637 ATCTGTGAGGAATCTGTCAAAATTCAGAAACCTATGCCGCAAAATAGCATPACAC 696
QY 687 GGCTGAAAAAATATGCAACGAGTCAGTTAAGCTATTCTTCAACGCTAGTACAA--- 743
Db 697 TTCAGAACAAACTGGGTACTCTTGTAAGCAATCCAGAACTACAAATATCAACAG 756
QY 744 -----TAAACGCAATCTGT 758
Db 757 CAACAACAGCAACACTAAGTCAAGCAAGTCAAAAGTATGATGATGATCTCTTGA 816
QY 759 AGCAAAAGGATCAACTAGCAAGCCAGCAAAATTAATCTGAAATCTCCAGATCTTTTGA 818
Db 817 ACAGCTTACAAACTGCTTGAAGTCAACGACATGTAATCTGATGCTGCTTCTTGA 876
QY 819 GGAACCTATATGATCACTAGCGCCCAAGTTACAGTGAATCAGATGCGCTGCTTGA 878
Db 877 TCCACAGCAAAATCAAGTCAAGGCTAGAGTGTGAGTCCACAGAGATCATTA 936
QY 879 CCTCTATGATTTATCAGTCAACCAAAATGAGTGGATTCGCAATGCGACCATTA 938
Db 937 CCACCTTATCTTCTTACAGCAAGCTTCTGCTTTAGAAAGAAAGATTTGCCAGAAATGCTGC 998
QY 999 TATCAGT 1005

```

RESULT 32
 AA65735
 ID AA65735 standard; DNA; 5048 BP.
 XX

```

AC AAA65735;
XX
DT 21-NOV-2000. (first entry)
XX
DE Streptococcus pneumoniae BVH-3 gene SEQ ID NO:11.
XX
KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal; ds.
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
DR WPI; 2000-452397/39.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia.
XX
PS Example 6; Fig 14; 106pp; English.
XX
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence encodes the
CC S. pneumoniae BVH-3 protein antigen.
XX
SQ Sequence 5048 BP; 1709 A; 907 C; 1104 G; 1328 T; 0 other;
XX
Query Match 16.1%; Score 385.4; DB 21; Length 5048;
Beet Local Similarity 65.3%; Pred. No. 9.5e-87;
Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;
QY 37 TAAGAAAATAATCTGTTTCTTATATAGTGAACAAAGGACGCAAAAAACGAGAA 96
Db 1875 TAAGACAATAATCTGTTCTTATGTGATGGACCCAGTCAAGTCAAGAAAGTGAAA 1934
QY 97 TTGACTCCTGATGAGTTAGCAAGGTAAGGAATCAATGCTGAGCAAAATGTCATCA 156
Db 1935 CTTGACACCAAGCCAGGTTAGCCAGAAAAGAAATTCAGGCTGAGCAAAATGTTATCA 1994
QY 157 GATAACAGACCAAGGCTATGCTCATCTCAATGCGCAGCACTATCTTATTAAGTGA 216
Db 1995 AATTACAGATCAAGGCTATGTAACGTCACAGGTAACATATCAATTAATGGA 2054
QY 217 GGTCTCTATGACGCTATCATGATGAAGAAATTAATCAAGAAAGTCAAACTATAGCT 276
Db 2055 AGTCTCTATGATGCCCCCTTATGTAAGAAATCTGTAAGAAAGATCCAACTATCACT 2114
QY 277 AAAAGATGAGATTTTGTATATGAGTCAAGGTTGATGTTATCAAGTGAATGAAA 336
Db 2115 TAAAGACGCTGATATTTGTCAATGAAGTCAAGGTTGTTATCATCAAGTCAATGAAA 2174
QY 337 ATACTATGTTTACCTTAAGATGCTGCCAGGATATAGTCCGTACAAAAGAGAAAT 396
Db 2175 ATATTATGCTACCTGAAAGATGCACTCATGCTATATGTTCAACTTAAAGATGAAT 2234
QY 397 CAATGCAAAAAACAAGACATAGTCAACATGCTGTAAGTGAAGTCCAAAGAACGATG 456

```

```

Db      2235 CAATCGTCAAAAACAAGAACTGTCAAGATATGAGAAAGTTA-----ACTC 2282
Qy      457 TCGTGTTCCTTGGACGCTTGGCAAGACGCTATCTCAAGTATGTTATATCTTAA 516
Db      2283 TAAATGTTCTGTAGCAAGGTCTCAGGAGCATATGACAAATGATGTTATGTTTAA 2342
Qy      517 TCGTTCGATATTCATAGAGGATAGTACTGTGATGCTTATATCTTCTCAATGAGATCATA 576
Db      2343 TCCAGCTGATATTTATCGAAGATACGGGTAATGCTTATATCTTCTCAATGAGATCATA 2402
Qy      577 CCATTCATCTCTTAAGATGAGTATTCAGTACGAGTGGCTGCTGCAGAAAGCTTCTT 636
Db      2403 TCACTACATTCCTCAAAAGCGATTTATCTGATGTAATGAGCAGCATTAAGCCATCT 2462
Qy      637 ATCTGTGAGAGAAATCTGTCAAAATTCAGAAACCTATGCGCAGCAAAATAGCGATTAAC 696
Db      2463 GCGTGGAAAAATATGCAACGAGTCAAGTAACTTATCTTCAACAGCTAGTACAA--- 2519
Qy      697 TTCAAGAACAACTGGGTACTCTTGTGAAGCAATCCAGAACTACAAATACACACAG 756
Db      2520 -----TAACAGCAATCTGT 2534
Qy      757 CAACACACGCAACTAATACAGTCAAGATCAAGTAAATGATGATGATCTCTTGA 816
Db      2535 AGCAAAAGATCACTACCAAGCCGCAAAATTAATCTGAAAATCTCCAGAGCTTTTGA 2594
Qy      817 ACAGCTTCAAACTGCTTGTGATGACGACATGATGATGATGCTTGTCTTGA 876
Db      2595 GGAATCTATGATTCACCTAGGCGCCAGTACGTAATCAATGATGCTGCTTGA 2654
Qy      877 TCCAGCAAAATCAACAGTCAACAGTGAAGTGTGACAGTCCACAGAGATCATA 936
Db      2655 CCTGCTAAGATTAATCACTCGTACACCAAAATGAGATTCGATCCGATGGGACATTA 2714
Qy      937 CCATTCATCTCTTACTCTCAATGCTGAATGGAAGAAGATCGTCTTATTTCC 996
Db      2715 CCACCTTATCTCTTACAGCAAGCTTTCGTTTGAAGAAAAGATTCAGAAATGTGCC 2774
Qy      997 CCTGCT 1003
Db      2775 TATCAGT 2781

RESULT 33
ABK15101
ID      ABK15101 standard; DNA; 5048 BP.
AC      ABK15101;
XX      08-MAY-2002 (first entry)
DT      08-MAY-2002 (first entry)
DE      DNA encoding Streptococcus pneumoniae BVH-3 version #1.
KM      BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
        pneumonia; streptococcal bacterial infection; gene; ds.
OS      Streptococcus pneumoniae.
XX      Key      Location/Qualifiers
        CDS      1777..4896
                /*tag= a
                /product= "BVH-3"
                /note= "The gene is flanked by sequences from the
                vector Spe4, no information on which is
                given in the specification"

MO200198334-A2.
PN      27-DEC-2001.
PD      19-JUN-2001; 2001WO-CA00908.
PF      20-JUN-2000; 2000US-212683P.
PR

```

```

XX      (SHIR-) SHIRE BIOCHEM INC.
PA      Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX      WPI; 2002-122272/16.
XX      P-PSDB; AAU75932.
XX      New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX      epitope-bearing polypeptides, useful as vaccine components for treating
XX      or preventing streptococcal infections such as otitis media,
XX      meningitis, and bacteraemia.
XX      Example 3; Fig 2; 113pp; English.
XX      The invention describes an isolated polypeptide (I) with 70-90%
XX      identity to Streptococcus pneumonia protein BVH-3. BVH-11, variants of
XX      BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX      comprising (I) is useful for therapeutic or prophylactic treatment of
XX      meningitis, otitis media, bacteraemia or pneumonia infection in an
XX      individual susceptible to these disorders. (II) is also useful for
XX      therapeutic or prophylactic treatment of any streptococcal bacterial
XX      infection (e.g., caused by Streptococcus pneumoniae, group A
XX      Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
XX      as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novardi or
XX      Streptococcus aureus) in an individual susceptible to the infection.
XX      A polynucleotide (III) encoding (I) is useful in DNA immunisation
XX      techniques. The Streptococcus polypeptides are useful in a diagnostic
XX      test for S. pneumoniae infection. (III) is useful for designing DNA
XX      probes for use in detecting the presence of Streptococcus in a biological
XX      sample suspected of containing the bacteria. The DNA probes may also be
XX      used for detecting circulating S. pneumonia nucleic acid in a sample for
XX      diagnosing streptococcal infections. This sequence encodes the
XX      Streptococcus pneumoniae protein BVH-3, used to create the antigenic
XX      peptides described in the method of the invention.
XX      Sequence 5048 BP; 1709 A; 907 C; 1104 G; 1328 T; 0 other;
XX      Query Match      16.1%; Score 385.4; DB 24; Length 5048;
XX      Best Local Similarity 65.3%; Pred. No. 9.5e-87;
XX      Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;
Qy      37 TAAAGAAATTAATGCTGTTTCTTATATGATGAAAAACAAGCAAGCAAAACGAGAA 96
Db      1875 TAAAGCAATATATGTTCTTATGATGAGGACGACATCAAGTCAAGAAATGAAAA 1934
Qy      97 TTGACTCTGATGAGGTGACCAAGCGTGAAGATCAATGCTGACCAATCTCATCA 156
Db      1935 CTGACACACAGCCAGGTTAGCCGAAAGAAATTCAGGCTGAGCAAAATGTAATCAA 1994
Qy      157 GATACAGACCAAGGCTATGCACTTGCATGAGGACACATCATATTATCAATGTGA 216
Db      1995 AATTACAGATCAGGCTATGATGACGACGAGGACCACTATATCTATATGAGAA 2054
Qy      217 GGTTCCTTATGACGCTATCATAGTGAAGATTAATCATGAAGATCCAAATTAAGCT 276
Db      2055 AGTTCCTTATGATGCTCCTCTTATGTAAGAACTTGTATGAAGATCCAAATCACT 2114
Qy      277 TAAAGATGAGATATTTGTTATGAGTCAAGGCTGATGTTATCAAGATGATGAAA 336
Db      2115 TAAAGACGCTGATATTTGTTATGAGTCAAGGCTGATGTTATCAAGGCTGATGAAA 2174
Qy      337 ATACTATGTTTACCTTAAAGATGCTGCCACGCGGATTAAGTCTGCTCAAAAGAGAAAT 396
Db      2175 ATATTATGTTCTACTGAAAGATGAGTCAATGCTGATGATGTTGAACTTAAAGTGAAT 2234
Qy      397 CAATGCAAAAAACAAGAGATAGTCAACATCGTGAAGTGAACCTCAAGAAACGATGG 456
Db      2235 CAATCGTCAAAAAACAACAATGTCAAAGATATGAGAAGTTA-----ACTC 2282
Qy      457 TCGTGTTCCTTGGACGCTTGGCAAGACGCTATCTCAAGTATGTTATATCTTAA 516
Db      2283 TAAATGTTCTGTAGCAAGGTCTCAGGAGCATATGACAAATGATGTTATGTTTAA 2342

```

QY 517 TCGTTCGATATCATAGAGGATCTGATGCTTATTCCTTCATGAGATCATTA 576
 Db 2343 TCCAGCTGATATATGAGATACGGGTATGCTTATTCCTTCATGAGATCATTA 2402
 QY 577 CCATTACATTTCTTAAGATGATTCAGCTAGCGATGGCTGCTGCAGAACCTTCT 636
 Db 2403 TCACATCATTTCCCAAGGATTTATCTGTGTGATTAAGCAGCAGCTAAAGCATCT 2462
 QY 637 ATCTGTCGAGGAATCTGTCAATTCAGAACCTTATCCGCCCAAAATAGCATACAC 696
 Db 2463 GGTGGAATAAATATGCAACCGAGTCAGTTAGCTTATCTTCAACAGCTAGTACAA--- 2519
 QY 697 TTCAGAACAACTGGGATCTTCTGTGAACAATCCAGAACATCAAACTATACACAAG 756
 Db 2520 -----TACACGGAATCTGT 2534
 QY 757 CAACAACAGCAACATTAAGATCAAGCAAGTCAAAAGTAATGACATTTAGTCTTTGAA 816
 Db 2535 AGCAAAAGGATCAACTAGCAAGCCAGCAAAATTAATCTGAAATCTCAGAGTCTTTGAA 2594
 QY 817 ACAGCTCTACAACTGCTTTGAGTCAAGCAATGTAATCTGATGGCTTCTTTGA 876
 Db 2595 GGAACCTATGATTAACCTAGCGCCCAAGCTTACAGTGAATCAAGTGGCTGTCTTGA 2654
 QY 877 TCCAGCAAAATCACAAGTCAAGCTAGAGGTGTTGACAGTCCACAGAGATCATTA 936
 Db 2655 CCTGTGATGATTTATGAGTCAACCAATGAGTGGATTCGCCATGGCAGCATTA 2714
 QY 937 CCATTCATTCCTTACTCTCAATGTTCTGAATTGGAAGACGAATGCTCTTATTTCC 996
 Db 2715 CCACCTTATTTCTTACAGCAAGCTTCTGCTTGAAGAAAGATGCAAGATGGTGC 2774
 QY 997 CCTGT 1003
 Db 2775 TATCAGT 2781
 RESULT 34
 AAV52325/C
 ID AAV52325 standard; DNA; 6867 BP.
 XX AAV52325;
 AC 23-OCT-1998 (first entry)
 DT 23-OCT-1998 (first entry)
 XX Streptococcus pneumoniae genome fragment SEQ ID NO:192.
 DE Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
 KW Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
 XX computer readable medium; vaccine; pharmaceutical composition; de.
 OS Streptococcus pneumoniae.
 XX WO9818931-A2.
 PN 07-MAY-1998.
 PD 30-OCT-1997; 97WO-US19588.
 XX 31-OCT-1996; 96US-0029960.
 PR (HUMA-) HUMAN GENOME SCI INC.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX Bareish SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M;
 PI Kunsch CA, Rosen CA;
 XX MPI; 1998-27225/24.
 DR Computer-readable medium with recorded Streptococcus pneumoniae
 PT polynucleotide sequences - useful in diagnostic kits and assays, and
 PT pharmaceutical compositions and vaccines for Streptococcus
 XX pneumoniae

PS Claim 1; Page 1157-1161; 1409pp; English.
 XX The present invention describes a computer readable medium which has
 CC the nucleotide sequences SEQ ID NO:1 to 391 (AAV52134 to AAV52524)
 CC recorded on it, or a representative fragment or a sequence at least 95%
 CC identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
 CC SEQ ID NO:1 to 391 (AAV52134 to AAV52524) are genomic fragments from
 CC Streptococcus pneumoniae. The present invention also describes an
 CC isolated nucleic acid molecule encoding a homologue of any of the
 CC fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
 CC nucleic acid molecule is produced by a process comprising: (a) screening
 CC a genomic DNA library using as a probe a target sequence defined by any
 CC of the sequences in SEQ ID NO:1 to 391, identifying members of the
 CC library which contain sequences that hybridise to the target sequence and
 CC isolating the nucleic acid molecules from the members; or (b) isolating
 CC mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
 CC molecules whose nucleotide sequence is homologous to amplification
 CC primers derived from the fragment of the S. pneumoniae genome to prime
 CC the amplification and isolating the amplified sequences. The computer
 CC readable medium can be used in a computer-based system for identifying
 CC fragments of the S. pneumoniae genome of commercial importance, or
 CC expression modulating fragments of the S. pneumoniae genome. Products
 CC from the present invention can be used in diagnosis kits and assays, and
 CC pharmaceutical compositions and vaccines for S. pneumoniae.
 XX
 SQ Sequence 6867 BP; 1896 A; 1325 C; 1212 G; 2433 T; 1 other;
 Query Match 16.1%; Score 385.4; DB 19; Length 6867;
 Best Local Similarity 65.3%; Pred. No. 1.1e-86;
 Matches 611; Conservative 0; Mismatch 276; Indels 60; Gaps 2;
 QY 37 TAAGAAAAATATCGTGTCTTATATGATGAAACAAAGGAGCAAAAAAGGAGAA 96
 Db 6687 TAAGCAATATATCGTGTCTTATATGATGAAACAAAGGAGCAAAAAAGGAGAA 6628
 QY 97 TTGAATCTGATGAGTGTGAGTGAAGCAAGTGAAGTGAAGTGAAGTGAAGTGAAG 156
 Db 6627 CTGACACGACGACGAGTTAGCCAGAAAGAAAGTGAAGTGAAGTGAAGTGAAG 6568
 QY 157 GATTAACAGCAAGGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 216
 Db 6567 AATTACAGATCAAGGCTATGATGATGATGATGATGATGATGATGATGATGATGAT 6508
 QY 217 GGTCTTATGAGCGTATCATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 276
 Db 6507 AGTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6448
 QY 277 AAAAGATGAGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 336
 Db 6447 TAAAGACGCTGATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6388
 QY 337 ATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 396
 Db 6387 ATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6328
 QY 397 CAATGACAAAAAAGAGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 456
 Db 6327 CAATGACAAAAAAGAGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 6280
 QY 457 TCGTGTGCTTGGACGTTCCAGAGAGCGATATCAAGATGATGATGATGATGATGAT 516
 Db 6279 TATGTTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 6220
 QY 517 TCGTGTGATATCATAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 576
 Db 6219 TCCAGCTGATATATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 6160
 QY 577 CCATTACATTTCTTAAGATGATTCAGCTAGCGATGGCTGCTGCAGAACCTTCT 636
 Db 6159 TCACATCATTTCCCAAGGATTTATCTGTGTGATTAAGCAGCAGCTAAAGCATCT 6100
 QY 637 ATCTGTCGAGGAATCTGTCAATTCAGAACCTTATCCGCCCAAAATAGCATACAC 696

```

Db      6099 GGGTGGAAAAATATGACACGAGTCAGTTAAGTATTCTTCACAGCCTAGTACAA--- 6043
Oy      6377 TTCAAGAACAACTGGGTACCTTCTGTAGCAATCCAGGAAGTACAAATACTAACACAG 756
Db      6042 -----TAAACGCAATCTGT 6028
Oy      757 CAACACAGCAACACTTAACAGTCAAGCAAGTCAAAGTATGATGATGATCTTTGAA 816
Db      6027 AGCAAAAGATCAATAGCAAGCAGCAAAATTAATCTGAAATCTCCAGAGCTTTTGA 5968
Oy      817 ACAGCTTACAAACTGCTTGTGATCAACGACATGTAGATGTGCTGCTTTTGA 876
Db      5967 GGAAGCTTATGATTCAGTACGCGCCCAAGTTACAGTATGATGAGCTGCTTTTGA 5908
Oy      877 TCCGCAAAATTCACAGTCCGACGCTAGAGTGTGACAGTCCACAGAGATCATTA 936
Db      5907 CCTGCTAAGATTCACAGTCCGACGCAAAATGAGATTCGATTCGCAATGGGACCATTA 5848
Oy      937 CCACCTTCACTCCCTTACTCTCAAAATGTCTGAAATGGAGAACGATCGCTGATTTATCC 996
Db      5847 CCACCTTATCTCTTACAGCAAGCTTCTGCTTGAAGAAAAAAGATTGCCAAGATGTGCC 5788
Oy      997 CCTTCGT 1003
Db      5787 TATCAGT 5781

RESULT 35
AAVS2488
ID      AAVS2488 standard; DNA; 973 BP.
AC      AAVS2488;
XX      23-OCT-1998 (first entry)
DE      Streptococcus pneumoniae genome fragment SEQ ID NO:355.
XX      Streptococcus pneumoniae; S. pneumoniae; genome; diagnosis; assay;
KM      computer readable medium; vaccine; pharmaceutical composition; ds.
XX      Streptococcus pneumoniae.
OS      Streptococcus pneumoniae.
XX      MO9818931-A2.
XX      07-MAY-1998.
PD      30-OCT-1997; 97WO-US19588.
PF      31-OCT-1996; 96US-0029960.
PR      (HUMA-) HUMAN GENOME SCI INC.
XX      PA
XX      Barash SC, Choi GH, Dillon PJ, Dougherty BA, Fannon M,
PI      Kunsch CA, Rosen CA;
XX      MPI; 1998-272225/24.
XX      DR
XX      Computer-readable medium with recorded Streptococcus pneumoniae
PT      polynucleotide sequences - useful in diagnostic kits and assays, and
PT      pharmaceutical compositions and vaccines for Streptococcus
PT      pneumoniae
XX      Claim 1; Page 1369-1370; 1409pp; English.
XX      CC
XX      The present invention describes a computer readable medium which has
CC      the nucleotide sequences SEQ ID NO:1 to 391 (AAVS2134 to AAVS2524)
CC      recorded on it, or a representative fragment or a sequence at least 95%
CC      identical to SEQ ID NO: 1 to 391. The nucleotide sequences depicted in
CC      SEQ ID NO:1 to 391 (AAVS2134 to AAVS2524) are genomic fragments from
CC      Streptococcus pneumoniae. The present invention also describes an
CC      isolated nucleic acid molecule encoding a homologue of any of the
CC      fragments of the S.pneumoniae genome (SEQ ID NO:1 to 391) where the
CC      nucleic acid molecule is produced by a process comprising: (a) screening

```

```

CC      a genomic DNA library using as a probe a target sequence defined by any
CC      of the sequences in SEQ ID NO:1 to 391, identifying members of the
CC      library which contain sequences that hybridize to the target sequence and
CC      isolating the nucleic acid molecules from the members; or (b) isolating
CC      mRNA, DNA or cDNA produced from an organism, amplifying nucleic acid
CC      molecules whose nucleotide sequence is homologous to amplification
CC      primers derived from the fragment of the S. pneumoniae genome to prime
CC      the amplification and isolating the amplified sequences. The computer
CC      readable medium can be used in a computer-based system for identifying
CC      fragments of the S. pneumoniae genome of commercial importance, or
CC      expression modulating fragments of the S. pneumoniae genome. Products
CC      from the present invention can be used in diagnosis kits and assays, and
CC      pharmaceutical compositions and vaccines for S. pneumoniae.
XX      SQ
XX      Sequence 973 BP; 327 A; 178 C; 211 G; 257 T; 0 other;
XX      Query Match      16.0%; Score 381.4; DB 19; Length 973;
XX      Best Local Similarity 77.5%; Pred. No. 4,7e-86;
XX      Matches 492; Conservative 0; Mismatches 131; Indels 12; Gaps 2;
Oy      1 TTCTTAGAGTTGGAGCTGATTCAGCTAGAACGTTAAGGAAA---TATCGTGTTC 57
Db      157 TTCTTAGAGCTTGAGAGTTACCAAGCTGTCAGAGTAAGAAAGTCAATCGAATGTC 216
Oy      58 CTATATAGATGAAAAACAGCGACGCAAAAAACGAGAAATTTGACTCCTGATGAGTTAG 117
Db      217 TTATATAGATGATGATACAGCTGTCGCAAAAGCGCAAAACCTTGACACAGATGAAAGTACG 276
Oy      118 CAAAGCGTGAAGAAATCAATGCTGAGCAAAATGCTCATCAAGTAAACAGACCAAGCTATGT 177
Db      277 TAAAGGGAGGGAGATCAACGCCGCAAAATTTGTTATCAAGATTCAGATCAAGGTTATGT 336
Oy      178 CACTTCACATGGGACGACACTATCATTTTCAAAAGTAAAGTTCCTATGACGTTATCAT 237
Db      337 GACCTTCATGAGAACCATTAATCAATCTATTAATGGAAGGTCCTTATGTCATCAT 396
Oy      238 CAGTGAAGAAATTAATCAATGAAAGATCCAAACTATTAAGCTAAAGATGAGATATTGTTAA 297
Db      397 CAGTGAAGAGCTCCTCATGAAAGATCCGAATTAATCAATGAGATTCAGACATTTGCA 456
Oy      298 TGAGGTCAAGGGTGAATGTTATCAAGTATGATGAAATTAATCAATGTTTACCTTTAAGA 357
Db      457 TGAATCAAGGGTGTGTTATGCTAATTAAGTAAACGTTAAATCAATGTTTACCTTTAAGA 516
Oy      358 TGCCTGCCACGCGGATTAACGCTGACAAAGAGAAATCAATGACAAACAAAGACA 417
Db      517 TGCAGCTCATGCGGATTAATTCGACAAAGAAAGAAATTAAGCTCAAGACGAAACG 576
Oy      418 TAGTCAACATCGTGAAGTGAATCCAGAAAGATGATGCTGCTTGGACAGCTTC 477
Db      577 CAGTCATATCAT-----AACTCAAGGACGATTAATGCTGTGACGACAGAGC 627
Oy      478 GCAAGACGCTATTAATCAAGATGATGTTATATCTTTAATGCTTGTATATCATAGAGA 537
Db      628 CCAAGAGAGTTATTAACAGCGATGATGATATCTTCAATGATCAATCATATCATTTGAGA 687
Oy      538 TACTGATGATGCTATATATGCTTCTCATGAGATCATTAACATTAATCTTAAGATGA 597
Db      688 CACGGGATGATGCTTATATGCTTCTCAAGGAGACATTAACATTAATCTTAAGATGA 747
Oy      598 GTTATCAGTAGAGAGTTGGCTGCTGAGAGCT 632
Db      748 GTTATCAGTAGAGAGTTAGCTGTGAGAAAGCT 782

RESULT 36
AAAS738
ID      AAAS738 standard; DNA; 2528 BP.
XX      AAAS738;
XX      AC
XX      21-NOV-2000 (first entry)
XX      DT
XX

```


CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating *S. pneumoniae* nucleic acid in a sample from
CC diagnosing streptococcal infections. This sequence encodes the
CC Streptococcus pneumoniae protein BVN-3, used to create the antigenic
CC peptides described in the method of the invention.

SQ Sequence 2528 BP; 960 A; 398 C; 505 G; 665 T; 0 other;

Query Match	Score	DB	Length
14.5%	347	24	2528

Matches 465; Conservative 0; Mismatches 160; Indels 12; Gaps 1;

OY	37	TAAGGAAAAATATCGGTTTCCATATATGATGAAAAACAAGCAGCAACCCAAAAAAGGAGAA	96
Db	39	TAGGACAAATATTCGTCTCTTATGTGATGCGACCACTCAAGTCAGAAAAGTAAAA	98
OY	97	TTTGACTCCTGATGAGGTAGCAAGCGGTGAAGGAATCAATGCTGAGCAAAATCGTCACA	156
Db	99	CTTGACACCCAGACCAGGTAGCCAGAAAGAGAAATTACAGGCTGACCAAAATTGTAATCAA	158
OY	157	GATACACACCAAGGCTATGTCACTTCATATGGCCGACACTATCATTTATTTACAATGTTAA	216
Db	159	AATTACAGATCGGGCTATGTAAACGTCACACGGTGATCACTATCATTTACTATATATGAGAA	218
OY	217	GGTTCCTTATGACGCTATCATCAGTAAAGAAATTACTCATGAAGATCCAAACTATTAAGCT	276
Db	219	AGTTCCTTATGATGCCCCCTTTAGTGAAGAACCTTTGATGAAGATCCAAACTATCAACT	278
OY	277	AAAAGATGAGATTTGTTTAATGAGGTCAGAGGTGATATGTTATCAAGGTAGATGAAA	336
Db	279	TAAAGACGTGTATTTGTCAATGAATCAAGGGTGTTATATCATCAAGGTGCATGAAA	338
OY	337	ATACTATGTTTACCTTAAAGATGCTGCCCAACCGGATACGTCCTGATCAAAAAGAGAAAT	396
Db	339	ATATTATGCTACCTGAAGAATGCAAGCTCATGCTGATTAATGTTTCAACTTAAGATGAAT	398
OY	397	CAATGCACAAAAACAAGACATAGTCAAGATCGTGAAGTGAACCTCCAAGAAACGATGG	456
Db	399	CAATCGTCAAAAACAAGACATGTCAAGATTAAGAAAGGTTAACTTAA-----	449
OY	457	TGCTGTGCTCTTGCGCACGTTCCGAGAGAGCGTCTACTACAGATGATGGTTATATCTTTAA	516
Db	450	---TGTTGTGTGAGCAAGGCTCTCAGGAGAGATTAAGACAAAGATGGTTATGTCTTTAA	506
OY	517	TGCTTCTGATATCATAGAGATACTGGTATGCTTATATCGTTCCTCATGAGATCATTTA	576
Db	507	TCCAGCTGATATTAATCGAAGATACGGGTAATGCTTATATCGTTCCTCATGAGAGTCACTA	566
OY	577	CCATTACATTCCTAAGAAATGAGTTATCAAGTCAGAGTTGGCTGCTGCAGAAACCTTCCCT	636
Db	567	TCACTACATTTCCCAAAAGGATTTATCTGCTAGTGAATTGACGACAGCTTAAACACATCT	626
OY	637	ATCTGGTCGAGGAATCTGTCAAAATTCMAAGACTTAT 673	
Db	627	GGCTGAAAAAAATATGCAACCGAGTACGTTAAACCTAT 663	

RESULT 38

ID AAA08556 standard; DNA; 504 BP.

AC AAA08556;

DT 19-JUL-2000 (first entry)

DE S. pneumoniae 20 kDa human C3-degrading protein coding sequence.

KW Human C3-degrading protein; 20 kDa; immunostimulatory; vaccine;

KW inhibitor; inflammation; organ rejection; xenotransplantation; ss.

Streptococcus pneumoniae.

PN WO200017370-A1.

XX 30-MAR-2000.
PD
XX
XX 24-SEP-1999; 99WO-US22362.
PF
XX
XX 24-SEP-1998; 98US-0101736.
PR
PR 31-MAR-1999; 99US-0283094.
XX
XX (MINU) UNIV MINNESOTA.
PA (AMCY) AMERICAN CYANAMID CO.
XX
XX
PI Hostetter MK, Finkel DJ, Cheng Q, Green BA, Masi AW;
XX
XX WPI; 2000-283594/24.
DR P-PSDB; AAY91938.
DR
XX
XX Isolated polypeptide is used to stimulate immune system and immunize or
PT treat a mammalian subject against Streptococcus pneumoniae infection or
PT colonization
XX
XX
PS Claim 1; Page 53; 61pp; English.

The present sequence, isolated from *Streptococcus pneumoniae*, encodes a human C3-degrading protein (see AA191938) of about 20 kDa. This C3 sequence may be part of a larger open reading frame (see AA08557) which encodes an approximately 92 kDa protein also having human C3-degrading activity. The DNA sequences can be used for producing an immune response to *Streptococcus pneumoniae* in a mammal. Antibodies against the C3 proteins can be used to inhibit *S. pneumoniae*-mediated C3 degradation. C3-mediated inflammation and rejection in xenotransplantation can be inhibited by expressing the nucleic acid sequences on the surface of an organ of an animal. In particular, the polypeptides are useful for stimulating the immune system and are effective to immunize or treat a mammalian subject against *Streptococcus pneumoniae* infection or colonization.

Sequence 504 BP; 153 A; 125 C; 98 G; 128 T; 0 other;

Query Match	10.48;	Score 247.4;	DB 21;	Length 504;
-------------	--------	--------------	--------	-------------

Matches 333; Conservative 0; Mismatches 106; Indels 12; Gaps 1;

QY	833	UCCTTGAAGTCAAGCAACATGTATAGATTCGTATAGGCTCTTGTCTTTGATTCAGACCAAAATGACA	892
Db	34	CCCTTATCAGAAACCCCATGTATAGATTCGTATAGGCTCTTATTTTCAGCCAGCGCAAAATACACA	93
QY	893	AGTGAACAGCTAAGAGTGTTCAGTGCACACGAGATCATTTACACTTCATCCCTTAC	952
Db	94	AGTGAACCGGCATATGGTGTTCGTATACCGACGAGAACCATTTATCACTTTATTCCTTAT	153
QY	953	TCTCAAAATGTCGAATTGGAAAGACGATGCGTCGTATTTATTTCCCTTCCTTATTCGTTCA	1012
Db	154	TCACAACCTGTACCTTTGGAAAGAAAATTTGGCTGTATTTATTCCTTCGTATTCGTCCA	213
QY	1013	AACCATTTGGGTACAGATTTCAAAGGCCAAGAACCAAGTCCACACACGACTCCGGAACCT	1072
Db	214	AACCATTTGGGTACAGATTTCAAGCCAGAACCAACCAAGTCCACATATGACTCCGGAACCT	273
QY	1073	AGTCCAGGCCCGCAACCTGCACCAAAATCTTAAATATAGACTCAA-----ATTCT	1120
Db	274	AGTCCAAAGTCCGCAACCTGCACCAAAATCTCTCAACCACTCCCAACCAATTCCAATTGATGAG	333
QY	1121	TCTTTTGGTGAATCAGCTGTATCCGAAAGATTGGGGAAGATATGTATTCGAAAGAAAGGCG	1180
Db	334	AAATTTGGTCAAGAGAGCTGTTCGAAAGATAGGCGATGTATATGCTTTTGGAGAAATGGA	393
QY	1181	ATCTCTCGTTATGTCTTTTGGCAAGAAATTACATCTCAAACTGTTAAAAATCTTGAAGC	1240
Db	394	GTTCTCTGTTATATATCCAGCCAAAGATCTTTCAACAGAAACAGCAGCAGGCATTGATGAC	453
QY	1241	AAATTATCAAAACAGAGAGTGTTCACACA	1271
Db	454	AAACTGGCCAAAGCAGAAAGTTTATCTCATTA	484

CC production of antibodies for use in passive immunotherapy, for use as
CC diagnostic reagents and for use as reagents in other processes such as
CC affinity chromatography.

XX Sequence 2478 BP; 846 A; 486 C; 518 G; 628 T; 0 other;

Query Match 9.0%; Score 214.6; DB 22; Length 2478;
Best Local Similarity 60.8%; Pred. No. 9.4e-44;
Matches 402; Conservative 0; Mismatches 229; Indels 30; Gaps 2;

```

OY 4 TTACAGAGTTGGACGCTGATCAAGTAGAAGCGTTAAGAAATATCGGTTCTTATAT 63
DB 66 TTACCAACTGGTAAAGCATATATGAGTTGACCAACAAAGGACATCAATTCCTATAT 125
OY 64 AGATGAAAACAAAGCGACGCAAAAA-----ACGGAATTTGACTCCTGATGAGTTAG 117
DB 126 TGATATATGCAAGGTAAGGCAAAAGCCCTTAAAAACAAACAAAGATGATCAATCAG 185
OY 118 CAAGCGTGAAGGAATCAATGCTGAGCAAAATCGTATCAAGATTAAGACCAAGGCTATGT 177
DB 186 TGCTGAAGAGGATCTCTGCTGAAACAGATCGTATCAAAATTTACTGACCAAGGCTATGT 245
OY 178 CACTTCACATGGCGACCACTATCATTTATACAAATGTTAGGTTCTTATGACGCTATCAT 237
DB 246 GACCTCACATGGTGAACCATTTATCAATTTTAAATGGAAGGTTCCCTTATGATGCAATAT 305
OY 238 CAGTGAAGATTTACTATCAAGAAATCCAACTATTAAGCTTAAAGATGAGATATTGTTAA 297
DB 306 TAGTGAAGAGTTGTTGATGACGATCTTAATTAACGTTTAAACATCAGAGTTATCA 365
OY 288 TGAGGTCAAGGTTGATATGTTATCAAGGTAGATGAAAAATATGTTTACCTTAAGGA 357
DB 366 TGAATCTTAGACGGTTACGTTATTAAGTCAATGGAACCTATATTGTTTACCTCAAGCC 425
OY 358 TGCTCCCAACGCGGATTAAGTCCCTGATCAAAAGAAATCAATCGACAAAAACAAGAGA 417
DB 426 AGGTAGCAAGCGCAAAACATTCGAACCAACAAATTTGCTGAGCAAGTGCAGAAAG 485
OY 418 TAGTCAACATCGTGAAG-----GTGAACTCCAAAGAAACGA 453
DB 486 AACTAAAGAGCTAAAGAAAAAGTTTAACTCAATGAGGCCCATCTCAGTAAGAAAGACT 545
OY 454 TGGTCTGTGGCTTGGCACTTCCAAAGACGCTATACATGATGATGTTATATCTT 513
DB 546 TGGCGAGTCAATGAAGCAAAAGAACAGAGAGCTATATACAGAGATGCTATATTTT 605
OY 514 TAACTCTTGTATCATAGAGATACGTGATGCTTATATGCTTCTCATGAGATCA 573
DB 606 TAGTCCGACAGATATCATATGATGATTTAGAGATGCTTATTTAGTACCTCATGATCA 665
OY 574 TTACATTACATTCCTAAGAAATGAGTTATCAGCTAGCGATGGCTGCGAGAGCCTT 633
DB 666 CTATATATATATCTTAAAGAAAGATTTGTTCCAAAGTAGAGTACGCTGCAAGACCTTA 725
OY 634 C 634
DB 726 C 726

```

RESULT 42

AAA05814 ID AAA05814 standard; DNA; 1146 BP.

XX AC AAA05814;

XX DT 30-MAY-2000 (first entry)

XX DE Group B Streptococcus protein encoding nucleotide sequence SEQ ID NO:12.

XX XX Group B Streptococcus; Streptococcus agalactiae; protein antigen;

KM vaccine; screening; immunogen; detection; diagnosis; infection;

XX antibody; antibody; antibacterial; ds.

OS Streptococcus agalactiae.

XX MO200006736-A2.

XX PN 10-FEB-2000.

XX PF 27-JUL-1999; 99WO-GB02444.

XX PR 27-JUL-1998; 98GB-0016335.

XX PR 19-MAR-1999; 98US-0125163.

XX (MCR-) MICROBIAL TECHNIQS LTD.

PI Le Page RMF, Wells JM, Hanniffy SB;

PT MPI: 2000-195299/17.

DR P-PSDB; AAY91286.

XX New Group B Streptococcus protein, useful as vaccine, for diagnosis of

XX Streptococcal infections and for screening of antibodies or affibodies

XX Claim 4; Fig 1; 123pp; English.

XX AAA05803 to AAA05872 encode proteins, polypeptides and peptides (given

XX in AAY91275 to AAY91343) isolated from Group B Streptococcus (GBS), also

XX known as Streptococcus agalactiae. The GBS polynucleotides and

XX polypeptides have antibacterial activity. Immunogenic compositions

XX comprising GBS polynucleotides or polypeptides can be used as vaccines

XX and for the treatment or prophylaxis of GBS infection. The

XX polynucleotides and polypeptides can also be used in the detection of GBS

XX and for screening DNA encoding bacterial cell envelope associated or

XX secreted antigens in gram positive bacteria. AAA05873 to AAA05941

XX represent primers used in the exemplification of the present invention.

XX Sequence 1146 BP; 389 A; 229 C; 234 G; 294 T; 0 other;

Query Match 8.8%; Score 211.4; DB 21; Length 1146;
Best Local Similarity 60.5%; Pred. No. 4.3e-43;
Matches 400; Conservative 0; Mismatches 231; Indels 30; Gaps 2;

```

OY 4 TTACAGAGTTGGACGCTGATCAAGTAGAAGCGTTAAGAAATATCGGTTCTTATAT 63
DB 66 TTACCAACTGGTAAAGCATATATGAGTTGACCAACAAAGGACATCAATTCCTATAT 125
OY 64 AGATGAAAACAAAGCGACGCAAAAA-----ACGGAATTTGACTCCTGATGAGTTAG 117
DB 126 TGATATATGCAAGGTAAGGCAAAAGCCCTTAAACAAACAAAGATGATCAATCAG 185
OY 118 CAAGCGTGAAGGAATCAATGCTGAGCAAAATGCTATCAAGATTAACAGACCAAGGCTATGT 177
DB 186 TGCTGAAGAGGATCTCTGCTGAAACAGATGCTAATTAATTAAGTCAAGAGTTATGT 245
OY 178 CACTTCACATGGCGACCACTATCATTTATTAACATGTAAGGTTCTTATGAGCTATCAT 237
DB 246 TACCTCACACGCTGACATATATTTTAAATGGAAGGAAATGCTTATGATGAGATATAT 305
OY 238 CAGTGAAGATTTACTATCAAGAAATGCAATGCTATTAAGTAAAGATGAGATATTGTTAA 297
DB 306 TAGTGAAGAGTTGTTGATGACGATCTTAATTTTAAACATGCAATGATGATCA 365
OY 298 TGAGGTCAAGGTTGATATGTTATCAAGGTAGATGGAATAATGTTTACCTTAAGGA 357
DB 366 TGAATCTTAAACGTTACGTTATTAATTAAGTCAATGCAATGATGTTTACCTCAAGCC 425
OY 358 TGCTGCCACGCGGATTAAGTCCGTTCAAAAGAGAAATCAATGCAAAAAACAAGACA 417
DB 426 AGGTAGTAAGGCAAAAAACATTCGAACCAACAAATTCCTGAGCAAGTACCAAGG 485
OY 418 TAGTCAACATGCTGAG-----GTGAACTCCAAAGAAAGA 453
DB 486 AACTAAAGAACTTAAAGAAAAAGTTTAACTCAAGTGGCCCATCTCAGTAAAGAAAGT 545
OY 454 TGGTCTGTGGCTTGGCACTTCCAAAGACGCTATTAAGTATGATGATGTTATATCTT 513

```

```

Db      546  TCGCGCACTCATGAAGCAAAAAGACAGACGCTATCTACAGACGATGGCTATATTT 605
Oy      514  TAAATCTTCTGATATCATAGAGACTGCTGATCTTATATCTTCTCATGAGATCA 573
Db      606  TAGTCCGACAGATATCATATGATTTTGGAGATCTTATTTAGTACTCATGTGAAATCA 665
Oy      574  TTACCATTAACATTCCTTAAGAATGAGTTATCAGCTAGCGAGTTGGCTGTGCAAGACCTT 633
Db      666  CTATCATTTATATCTCTTAAAAAAGATTGTCTCCAAGTAGAGCTAGCTGCTGCACAAAGCTTA 725
Oy      634  C 634
Db      726  C 726

RESULT 43
ABN69535
ID      ABN69535 standard; DNA; 2466 BP.
XX
AC      ABN69535;
XX
DT      01-JUN-2002 (first entry)
XX
DE      Streptococcus polynucleotide SEQ ID NO 6983.
XX
KM      Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;
KM      group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KM      antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS      Streptococcus agalactiae.
XX
PN      WO200234771-A2.
XX
PD      02-MAY-2002.
XX
PF      29-OCT-2001; 2001WO-GB04789.
XX
PR      27-OCT-2000; 2000GB-0026333.
PR      24-NOV-2000; 2000GB-0028727.
PR      07-MAR-2001; 2001GB-0005640.
XX
PA      (CHIR-) CHIRON SPA.
PA      (GENO-) INST GENOMIC RES.
XX
PI      Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI      Tectelin H;
XX
DR      WPI; 2002-352536/38.
DR      P-PSDB; ABP28904.
XX
PT      New Streptococcus protein for the treatment or prevention of infection
PT      or disease caused by Streptococcus bacteria, such as meningitis, and
PT      for detecting a compound that binds to the protein -
XX
PS      Claim 7; Page 3860-3861; 4525pp; English.
XX
CC      The invention relates to a protein (ABP25413-ABP30895) from group B
CC      streptococcus/GAS (Streptococcus agalactiae) or group A streptococcus/GAS
CC      (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC      the specification. The proteins have antibacterial and antiinflammatory
CC      activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC      antibodies that bind (I) are used in the manufacture of medicaments for
CC      the treatment or prevention of infection or disease caused by
CC      Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC      Nucleic acids encoding (I) are used to detect Streptococcus in a
CC      biological sample. (I) is used to determine whether a compound binds to
CC      (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC      used as a vaccine or diagnostic composition. The disease caused by
CC      Streptococcus that is prevented or treated may be meningitis. Nucleic
CC      acid encoding (I) may be used to recombinantly produce (I) and may be
CC      used in gene therapy. Antibodies to (I) are used for affinity
CC      chromatography, immunoassays, and distinguishing/identifying

```

```

CC      Streptococcus proteins.
XX
SQ      Sequence 2466 BP; 841 A; 475 C; 513 G; 637 T; 0 other;
Query Match      8.8%; Score 211.4; DB 24; Length 2466;
Best Local Similarity 60.5%; Pred. No. 6e-43;
Matches 400; Conservative 0; Mismatches 211; Indels 30; Gaps 2;

Oy      4  TTACAGTTGGGAGCTGATATCAAGCTAAGCGTTAAGGAAATATATGTTGCTTAT 63
Db      66  TTACAGCTTGGTAAAGATCATATGAGTGTCTAGACAAAGACATCAGATTGCTATAT 125
Oy      64  AGATGAAAAAAGCGACGAGAAAAA-----ACGAGAAATTGACTCCTGATGAGTTAG 117
Db      126  TGATGATAGCAAGATAGATTAAGTAAAGCCCTTAAACAAACAAACATGATCAATCAG 185
Oy      118  CAAGCGTGAAGGAATCAATGCTGAGCAAAATGCTCATCAAGTAAACAGACCAAGCTATGT 177
Db      186  TGCTGAAGAAAGGCATCTCTGCTGAAACAGATCGTAGTAAATTACTGACCAAGGTTATGT 245
Oy      178  CACTTCAATGGGAGCACTATCATCTTATTAATGTAAGGTTGCTTATGACGCTATCAT 237
Db      246  TACCTCAACGCTGACCATATATCTTTTACAAATGGAGAAAGTTCTTATGATGCGATTAT 305
Oy      238  CAGTGAAGAAATTACTCATGAAGATCCAACTAATAGCTTAAAGATGAGATATTGTTAA 297
Db      306  TAGTGAAGAGTTGTTGATGACGATCCTTAATTAACATTTTAAACATCAGACGTTATCAA 365
Oy      298  TGAAGTCAAGGGTGAATATGTTTATCAAGTAGATGAAATATCTATGTTTACCTTAAAGA 357
Db      366  TGAATCTTAGACGTTACGTTATTAAGTCAATGGAACATATATGTTTACCTCAAGCC 425
Oy      358  TGCCTGCCACCGGATTAACGTCCTGACAAAGAGAAATCATGACAAAAACAAGCA 417
Db      426  AGGTAGTAAGCGCAAAAACATTGACACCAACAAATATGCTAGCAAGTAGCCAAAGG 485
Oy      418  TAGTCAACATCGTAGAG-----GTGAACTCCAGAAAAAGCA 453
Db      486  AACTTAAAGAAAGCTTAAAGAAAAAGTTTAGCTCAAGTGCCCATCTCAGTAAAGAAAGT 545
Oy      454  TGGTCTGTTGCTTGGACGTTTCGCAAGAGCGCTATATCAAGATATGTTATCTT 513
Db      546  TCGCGCAGTCAATGAACCAAAAAGACAAGACCGCTATATCAACACCATGCTTATATTT 605
Oy      514  TAAATCTTCTGATATCATAGAGTACTGAGATGCTTATATGTTCTCATGAGATCA 573
Db      606  TAGTCCAGACAGATATCATATGATGTTTAGAGAGATCTTATTTATACCTCATGTAATCA 665
Oy      574  TTACCATTAACATTCCTTAAGAATGAGTTATCAGCTAGCGAGTTGGCTGTGCAAGACCTT 633
Db      666  CTATCATTTATATCTCTTAAAAAAGATTGTCTCCAAGTAGAGCTAGCTGCTGCACAAAGCTTA 725
Oy      634  C 634
Db      726  C 726

RESULT 44
ABN70334
ID      ABN70334 standard; DNA; 2466 BP.
XX
AC      ABN70334;
XX
DT      01-JUN-2002 (first entry)
XX
DE      Streptococcus polynucleotide SEQ ID NO 8581.
XX
KM      Streptococcus; GAS; group B streptococcus; Streptococcus agalactiae;
KM      group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
KM      antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
XX
OS      Streptococcus agalactiae.

```

PN WO200234771-A2.
 XX 02-MAY-2002.
 PD
 XX
 PF 29-OCT-2001; 2001WO-GB04789.
 XX
 XX 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 PI Telford J, Maignani V, Margarit Ros Yi, Grandi G, Fraser C;
 PI Tetrelin H;
 DR
 DR MPI: 2002-352536/38.
 DR P-PSDB: ABP29703.
 PT
 PT New Streptococcus protein for the treatment or prevention of infection
 PT for detecting a compound that binds to the protein -
 XX
 PS Claim 7, Page 3969; 4525pp; English.
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC Streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
 CC the specification. The proteins have antibacterial and anti-inflammatory
 CC activity. (I), nucleic acids encoding (I), ABN6044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for
 CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 CC
 SQ Sequence 2466 BP; 841 A; 475 C; 513 G; 637 T; 0 other;
 Query Match 8.8%; Score 211.4; DB 24; Length 2466;
 Best Local Similarity 60.5%; Pred. No. 6e-43;
 Matches 400; Conservative 0; Mismatches 231; Indels 30; Gaps 2;
 QY 4 TTACGAGTTGGGAGCTGTATCAAGCTAGACGGTTAGGAAATTAATCGTTTCTTAT 63
 DB 66 TTACCGAGCTTTGTAACATCATATATGAGTCTAGCAACAAAGACAAATGCTTATAT 125
 QY 64 AGATGAAAAACAAGCAGCCAAAAA-----ACGGAATTTGACTCTGATGAGGTTAG 117
 DB 126 TGATGATACAAAGTAAGTAAGAAACCCCTAAACAAACAAAGCATGATCAATTCG 185
 QY 118 CAAAGCTGAGAGTAATGCTGAGCAATTCATCAAGATPAACAGCAAGGCTATAT 177
 DB 186 TGCTGAGAGGAGCATCTCTGTAACAGATCGTAGCAAAATTAAGCAAGGTTATGT 245
 QY 178 CACTGCACATGGGAGCACTATCATATTAACAATGTAAGTTCCTTAGAGGCTATCAT 237
 DB 246 TACCTACACGGGTGACATTAATTAATTTTACAAATGGAAAGTTCTTAATGATGCGATTAT 305
 QY 238 CAGTGAAGATTACTCATGAAGATCAACTATAAGTAAAGATGAGATATTTGTTAA 297
 DB 306 TAGTGAAGGTTGTTGATGATGACGATCTTAATTAACATTAACATGACGATATCA 365
 QY 298 TGAAGTCAAGGTTGATGATGATCAAGGTAAGTGAAGAAATCTATGTTTACCTTAAGA 357
 DB 366 TGAATATCTTAGACGGTTAGCTTATTAAGTCAATGAGCACTATTATGTTTACCTCAAGCC 425

QY 358 TGTGTCGCCACGGGATTAACGTCGTACAAAGAGAAATCAATCGACAAACAAAGAGCA 417
 DB 426 AGTGTAGTAAGCCCAAAACATTCGACCAACAAATTTGCTGAGCAAGTACCCAAAG 485
 QY 418 TAGTCAACATCGTGAAG-----GTGAACTCCAGAAACGA 453
 DB 486 AACTAAGAGAGCTAAAGAAAAAGTTTAGCTCAAGTGCCCATCTCAGTAAGAAAGAGT 545
 QY 454 TGTGTCGTGTCCTTGCCGACCTTGCCAGAGAGCTATTAACAGATGATGATTAATCTT 513
 DB 546 TGGCGGAGTCAATGAAGCAAAAAAGCAAGAGGCTATCTACAGACGATGCTTAATTTT 605
 QY 514 TATGCTTCTGATATCATAGAGATAGTATGATGCTTATATGCTTCATGAGAGATCA 573
 DB 606 TAGTCGACAGATATCATTTGATGATTTAGAGATGCTTATTTAGTACCTCATGTAATCA 665
 QY 574 TTACCATTAATCTTCTTAAGATGATTAACAGCTTACAGCTAGCGATTGCTGCAAGACCTT 633
 DB 666 CTATCATTAATATCTCTTAAAGAAATTTGTCTCAAGTGAAGTACTGCTGCAAGCCCTA 725
 QY 634 C 634
 DB 726 C 726

RESULT 45
 AAA05811
 ID AAA05811 standard; DNA; 2469 BP.
 XX
 AC AAA05811;
 XX
 DT 30-MAY-2000 (first entry)
 XX
 XX Group B Streptococcus protein encoding nucleotide sequence SEQ ID NO:9.
 DE
 XX Group B Streptococcus; Streptococcus agalactiae; protein antigen;
 KW vaccine; screening; immunogen; detection; diagnosis; infection;
 KW antibody; affibody; antibacterial; ds.
 XX
 OS Streptococcus agalactiae.
 XX
 PN WO200006736-A2.
 XX
 PD 10-FEB-2000.
 XX
 PF 27-JUL-1999; 99WO-GB02444.
 XX
 PR 27-JUL-1998; 98GB-0016335.
 PR 19-MAR-1999; 99US-0125163.
 XX
 PA (MICR-) MICROBIAL TECHINICS LTD.
 PI Le Page RWF, Wells JM, Hanniffy SB;
 PI MPI: 2000-195299/17.
 DR P-PSDB: AA91283.
 DR
 XX New Group B Streptococcus protein, useful as vaccine, for diagnosis of
 PT Streptococcal infections and for screening of antibodies or affibodies
 PS
 PS Claim 4; Fig 1; 123pp; English.
 XX
 CC AAA05803 to AAA05872 encode proteins, polypeptides and peptides (given
 CC in AA91275 to AA91343) isolated from Group B Streptococcus (GBS), also
 CC known as Streptococcus agalactiae. The GBS polynucleotides and
 CC polypeptides have antibacterial activity. Immunogenic compositions
 CC comprising GBS polynucleotides or polypeptides can be used as vaccines
 CC and for the treatment or prophylaxis of GBS infection. The
 CC polynucleotides and polypeptides can also be used in the detection of GBS
 CC and for screening DNA encoding bacterial cell envelope associated or
 CC secreted antigens in gram positive bacteria. AAA05873 to AAA05941
 CC represent primers used in the exemplification of the present invention.

SQ Sequence 2469 BP; 842 A; 477 C; 514 G; 634 T; 0 other;

Query Match 8.8%; Score 211.4; DB 21; Length 2469;

Best Local Similarity 60.5%; Pred. No. 6e-43; Indels 30; Gaps 2;

Matches 400; Conservative 0; Mismatches 231; Indels 30; Gaps 2;

```

OY 4 TTACAGTTGGAGCTGTATCAAGTAGAAGGTTAAGAAATAATCGTGTCTTATAT 63
DB 66 TTACAGCTTGGTAGATCATATAGGCTCTAGCAACAAAGACATCATGATGCTATAT 125
OY 64 AGATGAAAACAGCGACGCAAAA-----ACGAGAATTGACTCCTGATGAGTTAG 117
DB 126 TGATGATAGCAAGGTAAGTAAAGCCCTTAAACAAACAAAGATGATCAAAATGAG 185
OY 118 CAAGGCTAAGAGATCAATGCTGAGCAAAATGCTATCAAGATACAGCAAGGTTATGT 177
DB 186 TGCTGAAGAGGCACTCTCTGTAAGAGATCGTAGTCAAAATTAATCTGACCAAGTTATGT 245
OY 178 CACTTCACATGGCGACCACTATCATTTATTAACATGTAAGTTCCTTATGACGCTATCAT 237
DB 246 TACCTCACAGCGTAGCATATATCATTTTACAATGGAAAGTTCCTTATGATGCGATTAT 305
OY 238 CAGTGAAGAATTACTCATGAAGATCAAACTATTAAGCTAATAAGATGAGATATTGTTAA 297
DB 306 TAGTGAAAGATTGTTGATGACGAGATCTTAATTACATTTTAAACATCAGACGTTATCAA 365
OY 298 TGAGGTCAAGGTGATATGTTATCAAGGTAGATGGAATAATCTGTTACCTTAAGGA 357
DB 366 TGAATCTTGAAGCGTTAGCTTATTAAGTCAATGCAACTATTAATGTTTACTTAAAGCC 425
OY 358 TGCTGCCACGCGGATTAACGTCCTGTAACAAAGAGAAATCAATCGACAAACAAAGACA 417
DB 426 AGGTAGTAAGGCAAAAACATTCGAACCAAAACAAATTTGCTGACAAAGTAGCCAAAGG 485
OY 418 TAGTCAATCTGTGAAG-----GTGAACTCCACAAAGCA 453
DB 486 AACTTAAAGAGCTTAAAGAAAAGTTTAAAGCTCAAGTGGCCCATCTCAGTAAAGAAAGT 545
OY 454 TGGTGCCTGTGCTGGACGCTTGGCAAGGACGATATCAATGATGATGATATATCTT 513
DB 546 TGGGCACTGATGAAGCAAAAAGACAAAGACGCTATCTACAGACGATGCTATATATTT 605
OY 514 TAATGCTCTGATATCATAGAGATATGATGATCTTATATCGTTCTCATGAGATCA 573
DB 606 TAGTCCGACAGATATCATATGATGATTTAGAGATGCTTATTTAGTACCTCATAGATCA 665
OY 574 TTACATTACATCTCTTAAGATGAGTTATCAAGCTAGCAAGTGGCTGCTGACAAAGCTT 633
DB 666 CTATCATTAATATCTCTTAAAAAAGATTGCTCAAGTAGAGCTGCTGCAAGGCTTA 725
OY 634 C 634
DB 726 C 726

```

RESULT 46

AA00038

ID AA00038 standard; DNA; 2469 BP.

AA00038;

11-MAY-2001 (first entry)

Streptococcus agalactiae DNA encoding GBS36.

Streptococcus agalactiae DNA encoding GBS36.

Group B streptococci protein 36; GBS36; immunogen; vaccine;

antibody; necrotising fasciitis; scarlet fever; impetigo;

bacterial meningitis; otitis media; community-acquired pneumonia; ds.

Streptococcus agalactiae.

Key

Location/Qualifiers

1..2469

FT /tag= a

/product= "GBS36"

XX WO200114421-A1.

XX 01-MAR-2001.

XX 25-AUG-2000; 2000WO-US23417.

XX 25-AUG-1999; 99US-0150750.

XX (MEDI-) MEDIMUNE INC.

XX Koenig S, Heinrichs J, Johnson LS, Adamou JE;

XX WPI: 2001-211305/21.

XX P-PSDB; AAU00028; AAU00030.

PT New polypeptides obtained from group A or B streptococci, especially

PS Claim 10; Page 53-54; 62pp; English.

The sequence encodes Streptococcus agalactiae Group B Streptococci protein 36, GBS36. A recombinant cell producing GBS36, GBS36(2) or GBS36 is useful as a vaccine for vaccinating an animal, preferably a human against infection by a bacterial organism such as a streptococcal or streptococcal bacterium, and for treating a disease caused by group A streptococci, group B streptococci or Streptococcus aureus in an animal preferably a human. Vaccines and antibodies against the proteins of the invention are useful in prophylaxis and/or treatment of diseases such as necrotising fasciitis, scarlet fever, sepsis, impetigo, bacterial meningitis, otitis media, community-acquired pneumonia and many diseases of newborns. The proteins are also used as immunogens to stimulate the production of antibodies for use in passive immunotherapy, for use as diagnostic reagents and for use as reagents in other processes such as affinity chromatography.

SQ Sequence 2469 BP; 844 A; 477 C; 514 G; 634 T; 0 other;

Query Match 8.8%; Score 211.4; DB 22; Length 2469;

Best Local Similarity 60.5%; Pred. No. 6e-43; Indels 30; Gaps 2;

Matches 400; Conservative 0; Mismatches 231; Indels 30; Gaps 2;

```

OY 4 TTACAGTTGGAGCTGTATCAAGTAGAAGGTTAAGAAATAATCGTGTCTTATAT 63
DB 66 TTACAGCTTGGTAGATCATATAGGCTCTAGCAACAAAGACATCATGATGCTATAT 125
OY 64 AGATGAAAACAGCGACGCAAAA-----ACGAGAATTGACTCCTGATGAGTTAG 117
DB 126 TGATGATAGCAAGGTAAGTAAAGAGCCCTTAAACAAACAAACGATGATCAATAG 185
OY 118 CAAGGCTAAGAGATCAATGCTGAGCAAAATGCTATCAAGATACAGCAAGGTTATGT 177
DB 186 TGCTGAAGAGGCACTCTCTGTAAGAGATCGTAGTCAAAATTAATCTGACCAAGTTATGT 245
OY 178 CACTTCACATGGCGACCACTATCATTTATTAACATGTAAGTTCCTTATGACGCTATCAT 237
DB 246 TACCTCACAGCGTAGCATATATATTTTAAAGGAGAAAGTTCCTTATGATGAGATTAT 305
OY 238 CAGTGAAGAATTACTCATGAAGATCCAAACTATTAAGCTTAAAGATAGATATTGTTAA 297
DB 306 TAGTGAAAGATTGTTGATGACGATCTTAATTACATTTTAAACATCAGACGTTATCAA 365
OY 298 TGAGGTCAAGGTGATATGTTATCAAGGTAGATGGAATAATCATGTTTAACTTAAGGA 357
DB 366 TGAATCTTGAAGCGTTAGCTTATTAAGTCAATAGGCAACTATATATGTTACTCAAGCC 425
OY 358 TGCTGCCACGCGGATTAACGTCCTGTAACAAAGAGAAATCAATGACAAACAAAGACA 417
DB 426 AGGTAGTAGAGGCAAAAACATTCGAACCAACAAATTTGCTGAGCAAGTAGCCAAAGG 485

```

QY 418 TAGTCACATCTGTGAG-----GTGGAATCCAGAAACGA 453
 Db 486 AACTAAAGAGCTAAAGAAAGAGTTAGCTCAAGTGCCCATCTCAGTAAAGAAAGT 545
 QY 454 TGTGCTGTGCTTGGCAGCTTGCAGAGCCCTTACTACAGATGATGTTATCTT 513
 Db 546 TGGGGAGTCATAGCAAGCAAAAAGACAGCCCTACTACAGACGATGGCTATATTTT 605
 QY 514 TAATGCTTCTGATATCATAGAGATCTGTGATGCTTATATGCTTCATGAGATCA 573
 Db 606 TAGTCCGACAGATATCATGATGATTTAGAGATGCTTATTTAGTACCTCATGTAATCA 665
 QY 574 TTACCATTAACATTCCTAAGAAATGATTAAGCTAGCAGATTGGCTGCAGAACCTT 633
 Db 666 CTATCATTTATATCTTAAAGAAAGATTTGCTCCAGTGAAGCTAGCTGCTGCAACAGCTTA 725
 QY 634 C 634
 Db 726 C 726

RESULT 47
 ABN71527/C
 ID ABN71527 standard; DNA; 2155561 BP.
 AC ABN71527;
 XX
 DT 02-JUL-2002 (first entry)

Streptococcus polynucleotide SEQ ID NO 10967.
 DE Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KM group A streptococcus; Streptococcus pyogenes; antibacterial; gene;
 XX antiinflammatory; infection; vaccine; meningitis; gene therapy; ds.
 OS Streptococcus sp.
 XX
 PN MO200234771-A2.
 XX
 PD 02-MAY-2002.
 XX
 PF 29-OCT-2001; 2001WO-GB04789.
 XX
 PR 27-OCT-2000; 2000GB-0026333.
 XX
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Telford J, Maignant V, Margarit Ros YI, Grandi G, Fraser C;
 XX Tectelin H;
 PI
 WI; 2002-352536/38.
 XX
 DR
 XX
 PT New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 PS
 PS Claim 8; Page 4196-4488; 4525pp; English.

The invention relates to a protein (ABP25413-ABP30895) from group B streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in the specification. The proteins have antibacterial and antiinflammatory activity. (I), nucleic acids encoding (I), ABN65044-ABN71526 and antibodies that bind (I) are used in the manufacture of medicaments for the treatment or prevention of infection or disease caused by Streptococcus bacteria, particularly S. agalactiae and S. pyogenes. Nucleic acids encoding (I) are used to detect Streptococcus in a biological sample. (I) is used to determine whether a compound binds to (I). A composition comprising (I) or a nucleic acid encoding (I), may be used as a vaccine or diagnostic composition. The disease caused by

CC Streptococcus that is prevented or treated may be meningitis. Nucleic acid encoding (I) may be used to recombinantly produce (I) and may be used in gene therapy. Antibodies to (I) are used for affinity chromatography, immunoassays, and distinguishing/identifying Streptococcus proteins.
 CC
 XX
 SQ Sequence 2155561 BP; 695741 A; 385320 C; 382958 G; 691542 T; 0 other;
 Query Match 8.8%; Score 211.4; DB 24; Length 2155561;
 Best Local Similarity 60.5%; Pred. No. 1.2e-41;
 Matches 400; Conservative 0; Mismatches 231; Indels 30; Gaps 2;

QY 4 TTACGAGTTGGAGCTGTATCAAGCTAGAACGTTTAAAGAAATATCGTGTCTTATAT 63
 Db 1213286 TTACAGCTGGTGAAGCATATATGAGTCTAGCAACAAAGACATCGATTGCTATAT 1213227
 QY 64 AGATGAAAAACAGCGACGCAAAA-----ACGGAATTTGACTCCTGATGAGCTTAG 117
 Db 1213226 TGATGATAGCAAGGTAAGTAAAGCCCTTAAACAAACAAAGATGATCAATCAG 1213167
 QY 118 CAAGCGTAAGGAATCATATGCTGAGCAAAATGCTCATCAAGATTAACAGCAAGCTATGT 177
 Db 1213166 TGCTGAAGAGGCACTCTGCTGAACAGATCGTACAAATTAAGTGAACAGGTTATGT 1213107
 QY 178 CACTTCACATGGCGACCACTATCATTTATTAACAATGTTAGCTTATGACGCTATCAT 237
 Db 1213106 TACCTCACAGCGTGACCATTTATCATTTTACATGGAAGTTCCTTATGATGCGATTAT 1213047
 QY 238 CAGTGAAGATTACTCATGAAAGATCCAAACTATTAAGCTTAAAGATGAGATATTGTTAA 297
 Db 1213046 TAGTGAAGAGTTGTTGATGACGATCTTAATTAACATTTTAAACATCAGACGTTATCA 1212987
 QY 298 TGAAGTCAAGGTGATATGTTATCAAGTATGAAATATCTATGTTTACTTAAAGGA 357
 Db 1212986 TGAATCTTGAAGCGGTACGTTATTAAGTCAATGCAACTATTAATGTTTACTTAAAGCC 1212927
 QY 358 TCGTCCCAACGCGGATTAAGCGTCCGTACAAAGAGAAATCAATGCAACAAACAGAGA 417
 Db 1212926 AGGTAGTAAGCGCAAAACATTCGAACCAACAAATTTGCTGAGCAAGTGGCCAAAG 1212867
 QY 418 TAGTCACATCTGTGAG-----GTGGAATCCAGAAACGA 453
 Db 1212866 AACTAAAGAGCTAAAGAAAGAGTTTACTCAAGTGCCCATCTCAGTAAAGAAAGT 1212807
 QY 454 TGTGCTGTGCTTGGCAGCTTGCAGAGCCCTTACTACAGATGATGTTATCTT 513
 Db 1212806 TGGGCGAGTCATAGCAAGCAAAAAGACAGCCCTACTACAGACGATGGCTATATTTT 1212747
 QY 514 TAATGCTTCTGATATCATAGAGATCTGTGATGCTTATATGCTTCATGAGATCA 573
 Db 1212746 TAGTCCGACAGATATCATGATGATTTAGAGATGCTTATTTAGTACCTCATGTAATCA 1212687
 QY 574 TTACCATTAACATTCCTAAGAAATGATTAAGCTAGCAGATTGGCTGCAGAACCTT 633
 Db 1212686 CTATCATTTATATCTTAAAGAAAGATTTGCTCCAGTGAAGCTAGCTGCTGCAACAGCTTA 1212627
 QY 634 C 634
 Db 1212626 C 1212626

RESULT 48
 AAA65740
 ID AAA65740 standard; DNA; 2469 BP.
 AC AAA65740;
 XX
 DT 21-NOV-2000 (first entry)

Streptococcus pneumoniae GBS BVH-71 gene SEQ ID NO:80.
 DE Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KM Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KM prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;

KM otitis media; pneumonia; immunisation; bactericidal; ds.
 XX Streptococcus pneumoniae.
 OS WO200039299-A2.
 PN XX
 XX 06-JUL-2000.
 PD
 XX 20-DEC-1999; 99WO-CA01218.
 PF
 XX 23-DEC-1998; 98US-0113800.
 PR
 XX (BIOC-) BIOCHEM PHARMA INC.
 PA
 XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 PI
 XX WPI; 2000-452397/39.
 DR F-PSDB; AAB12746.
 XX
 XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia -
 XX
 XX Example 13; Fig 45; 106pp; English.
 PS
 XX The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence encodes the
 CC S. pneumoniae GBS BVH-71 protein antigen.
 XX
 SQ Sequence 2469 BP; 842 A; 481 C; 519 G; 627 T; 0 other;
 Query Match 8.8%; Score 209.8; DB 21; Length 2469;
 Best Local Similarity 60.4%; Pred. No. 1.5e-42;
 Matches 399; Conservative 0; Mismatches 232; Indels 30; Gaps 2;
 Oy 4 TTACGAGTTGGGACTGTATCAAGCTAGAAAGGTTAAGGAAATATCGTTTCTTAT 63
 Db 66 TTACCACTTGTGTAGCATCATATGAGTGTACCAAGGCAATCATGATGCTTATAT 125
 Oy 64 AGATGAAAGCAAGCGAGCAAAAA-----ACGAGATTGTCCTCGATGAGCTATAG 117
 Db 126 TGATACGCAAGGTTAAGGCAAAAGCCCTTAAACAAACAAAGATGATCAATCAG 185
 Oy 118 CAAGCGTGAAGGATCAATGCTGAGCAAAATCGTCATCAAGATACAGCAAGGCTATGT 177
 Db 186 TGCTGAAGAGGCACTCTGCTGAGACAGATCTGTAATAAATTACTGACCAAGCTATGT 245
 Oy 178 CACTTCACATGGCGCACTATCATTTATTACATGTAAGGTTCTTATGACGCTATCAT 237
 Db 246 GACCTCACAAGGTCACATTTATCAATTTTCAATGGGAAAGTTCTTATGATGCGATAT 305
 Oy 238 CAGTGAAGATTACTCATGAAGATCCAAACTATAGCTAAAGATGAGGATATGTTAA 297
 Db 306 TAGTGAAAGGTTGTGATGACGATCTTATTAACCGTTTAAACAATCAGACGTTATCA 365
 Oy 298 TGAGGTCAAGGTTGATATGTTATCAAGTAGATGAAATATCTATGTTTACTTAAGGA 357
 Db 366 TGAATCTTTAGAGCGTTACGTTATTAAGTCAATGCAACTTATTTAGTTTACTCAAGCC 425
 Oy 358 TGCTGCCACGCGATTAACGTCGTGTAACAAAGAGAAATCAATCGACAAAAACAAGACA 417
 Db 426 AGGTGTAAGCGCAAAACATTCGAACCAACAAATTGCTGAGCAAGTGGCCAAAG 485
 Oy 418 TAGTCAATCGTGAAG-----GTGAACCTCCCAAGAAAGA 453
 Db 486 AACTTAAGAGAGCTTAAGAAAAAGTTTACCTCAAGTGCCCAATCTCAGTAAAGAAAGT 545
 Oy 454 TGTGCTGTGTGCTTGACGCTTCCGAAGACGCTATACAGATGATGTTATATCTT 513

Db 546 TCGCGAGTCAATGAGCAAAAAAGCAAGACGCTATCTACAGACGATGCTATATTT 605
 Oy 514 TTAATCTTCATATCATAGAGTACGTGATGCTTATTCGTTCTCATGAGATCA 573
 Db 606 TAGTCCGACAGATATCATGATGATTTAGAGATGCTTATTTAGTACTCATGGTAATCA 665
 Oy 574 TTACCATTAACATTCCTAAGAAATGAGTTATCATGCTAGCGAGTTGGCTGTCAGAGCCTT 633
 Db 666 CTATCATTAATTTCTTAAAAAGATTGTCTCCCAAGTAGAGTACTGCTGCAAGACCTA 725
 Oy 634 C 634
 Db 726 C 726
 RESULT 49
 AAX91105
 ID AAX91105 standard; DNA; 5215 BP.
 XX
 XX AAX91105;
 AC
 XX 15-NOV-1999 (first entry)
 DT
 XX
 XX Group B Streptococcus (GBS) antigens encoding DNA (clone 3).
 DE
 XX
 XX Group B Streptococcus; GBS; antigen; vaccine; Streptococcus infection;
 KW sepsis; meningitis; pneumonia; immunocompromise; diabetes; liver disease;
 KW cancer; velinary; mastitis; se.
 XX
 OS Streptococcus sp.
 XX
 XX
 FH Key location/Qualifiers
 FT CDS 3..125
 FT /tag= a
 FT /note= "see AAY27346"
 FT 133..2514
 FT /tag= b
 FT /note= "see AAY27347"
 FT 367..2514
 FT /tag= c
 FT /note= "see AAY27348"
 FT complement (2716..2946)
 FT /tag= d
 FT /note= "see AAY27349"
 FT complement (2995..3252)
 FT /tag= e
 FT /note= "see AAY27350"
 FT complement (3299..3676)
 FT /tag= f
 FT /note= "see AAY27351"
 FT complement (3837..4124)
 FT /tag= g
 FT /note= "see AAY27352"
 FT complement (4351..5214)
 FT /tag= h
 FT /note= "see AAY27353"
 FT
 FN W09942588-A2.
 XX
 XX 26-AUG-1999.
 PD
 XX 17-FEB-1999; 99WO-CA00114.
 PE
 XX 20-FEB-1998; 98US-0075425.
 PR
 XX (BIOC-) BIOCHEM VACCINS INC.
 PA
 XX Boyer M, Brodeur BR, Charlebois I, Hamel J, Martin D;
 PI Rioux C;
 XX WPI; 1999-540309/45.
 DR P-PSDB; AAY27346, AAY27347, AAY27348, AAY27349, AAY27350, AAY27351,

DR AA127352, AA127353.
 XX Novel group B Streptococcus antigens - useful as vaccine
 PT compositions for prophylaxis or therapy of Streptococcus infections
 XX
 BS Claim 10; Fig 3A; 154pp; English.
 CC The invention provides Group B Streptococcus (GBS) antigens
 CC (AA127352-370) and nucleic acids (AA127352-370) encoding the antigens.
 CC The GBS antigens can be recombinantly expressed using standard
 CC recombinant methodology. The GBS antigens of the invention can be used as
 CC vaccine components for the treatment or prophylaxis of diseases and
 CC symptoms mediated by Streptococcus infection, especially group A
 CC Streptococcus (S. pyogenes), GBS or S. agalactiae, S. dysgalactiae, S.
 CC uberis, S. novae, as well as Streptococcus agalactiae. The vaccines are
 CC administered to those individuals at risk of GBS infection, particularly
 CC pregnant women and infants for sepsis, meningitis, and pneumonia, as well
 CC as immunocompromised individuals, such as those with diabetes, liver
 CC disease or cancer. The vaccines also have veterinary applications, such as
 CC for the treatment of mastitis in cattle. The present sequence represents
 CC a DNA encoding 8 GBS antigens of the invention.
 XX
 SQ Sequence 5215 BP; 1631 A; 957 C; 1038 G; 1589 T; 0 other;

Query Match 8.8%; Score 209.8; DB 20; Length 5215;
 Best Local Similarity 60.4%; Pred. No. 2.1e-42;
 Matches 399; Conservative 0; Mismatches 232; Indels 30; Gaps 2;
 Oy 4 TTACGAGTTGGAGCTGTATCAAGCTAGAACGCTTAAGCAAAATATCGTTCTTCTATAT 63
 Db 111 TTACCACTTGTGTAAGCATATATGAGTGTACCAACAAAGGCAATCATGATGCTTATAT 170
 Oy 64 AGATGAAACAAAGGAGCGCAAAAA-----ACGAGAAATTGACTCCTGATGAGGTAG 117
 Db 171 TGATACAGCAAGGTAAGGCAAAAGCCCTTAAACAAACAAAGAGATCAATCAACG 230
 Oy 118 CAAGGTGAAGGAATCAATGCTGAGCAATCTCTCAATCAAGATAGACCAAGGCTATGT 177
 Db 231 TGCTGAAGAGGATCTCTGCTGGAACAGATGCTAGTCAAAATATCTGACCAAGGCTATGT 290
 Oy 178 CACTTCAATGCGGACCACTATCTATTTATCAATGTAAGGTTCTTATGACGCTATCAT 237
 Db 291 GACCTCAACGCTGACCAATTTATCTATTTTACATGGAAGTTCTTATGAGCATTTAT 350
 Oy 238 CAGTGAAGATTTACTCATGAAGATCCAAACTATAGCTAAAGTAAGATGAGATATGTTAA 297
 Db 351 TAGTGAAGGTTGTTGATGAGCGATCTTATTAACGTTTAAACATCAAGAGCTATCA 410
 Oy 298 TGAGGTCAAGGTGATATGTTATCAAGGTAGATGAAATATCTATGTTTACCTTAAGA 357
 Db 411 TGAATCTTAGACGCTTACGTTATTAAGTCAATGCAATGCAATCTTATGTTTACCTCAAG 470
 Oy 358 TGCTGCCACCGCGATTAACGTCGTAACAAAGAGAAATCAATGACCAAAAAACAAGCA 417
 Db 471 AGGTGTAAGCGCAAAAAATTCGAACCAACAAATGCTGAGCAAGTACCAAGG 530
 Oy 418 TAGTGAACATGCTGAG-----CTGAACCTCCCAAGAAAGA 453
 Db 531 AACTTAAGAGCTTAAGAAAAAGTTTAACTCAAGTGGCCATCTCAGTAAAGAAAGT 590
 Oy 454 TGCTGCTGCTTGGCGACGTTCCGAAGAGCGCTATATCAAGATGATGTTATATCTT 513
 Db 591 TGGCGCACTCAATGAAGCAAAAGACAGAGCGCTATCTACAGAGATGCTATATTTT 650
 Oy 514 TAACTCTTCTGATATCATAGAGATATCTGATGATCTTATATCTTCTCATGAGATCA 573
 Db 651 TAGTCCGACAGATATCATATGATGATTTAGAGATGCTTATTTAGACCTCATGTAATCA 710
 Oy 574 TTACATATACATCTTCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 633
 Db 711 CTATATATATTTCTTAAAGAGATTTGCTCAAGTGAAGTGAAGTGAAGTGAAGTGAAG 770
 Oy 634 C 634

Db 771 C 771
 RESULT 50
 AAA65741
 ID AAA65741 standard; DNA; 2472 BP.
 AC AAA65741;
 DT 21-NOV-2000 (first entry)
 DE Streptococcus pneumoniae GAS BVH-71 gene SEQ ID NO:82.
 XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KM otitis media; pneumonia; immunisation; bactericidal; ds.
 XX Streptococcus pneumoniae.
 OS
 XX MO200039239-A2.
 PN
 XX 06-JUL-2000.
 PD
 XX 20-DEC-1999; 99WO-CA01218.
 PF
 XX 23-DEC-1998; 98US-0113800.
 PR
 XX (BIOC-) BIOCHEM PHARMA INC.
 PA
 XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 PI MPI: 2000-452397/39.
 DR P-PDB; AAB12747.
 XX
 PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia -
 PS Example 13; Fig 47; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence encodes the
 CC S. pneumoniae GAS BVH-71 protein antigen.
 XX
 SQ Sequence 2472 BP; 846 A; 480 C; 514 G; 632 T; 0 other;

Query Match 8.6%; Score 206.6; DB 21; Length 2472;
 Best Local Similarity 60.1%; Pred. No. 9.8e-42;
 Matches 397; Conservative 0; Mismatches 234; Indels 30; Gaps 2;
 Oy 4 TTACGAGTTGGAGCTGTATCAAGCTAGAACGCTTAAGCAAAATATCGTTTCTTATAT 63
 Db 66 TTACCACTTGTGTAAGCATATATGAGTGTCTAGCAACAAAGGCAATCATGATGCTTATAT 125
 Oy 64 AGATGAAACAAAGGAGCGCAAAAA-----ACGAGAAATTGACTCCTGATGAGGTAG 117
 Db 126 TGATGATAGCAAGGTAAGGCAAAAGCCCTTAAACAAACAAAGATGATCAATCAAG 185
 Oy 118 CAAGGTGAAGGAATCAATGCTGAGCAAAATGCTCATCAAGATTAACAGACCAAGGCTATGT 177
 Db 186 TGCTGAAGAGGATCTCTGCTGAACAGATGCTGATCAAAATTTACTGACCAAGTATGT 245
 Oy 178 CACTTCAATGCGGACCACTATCTATTTATCAATGTAAGGTTCTTATGACGCTATCAT 237
 Db 246 GACCTCAACGCTGACCAATTTATCTATTTTAAATGGAAGATTTCTTATGATGCAATTTAT 305
 Oy 238 CAGTGAAGATTTACTCATGAAGATCCAACTATATAGCTAAAGATGAGATATTTGTTAA 297

```
Db 306 TAGTAAAGAGTTGTGATGACGATCCTAATTACCATTTTAAACAATCAGAGTTATCAA 365
Oy 298 TGAAGTCAAGGTGGATATGTTATCAAGTAGATGGAATAATCTATGTTTACCTTAAGGA 357
Db 366 TGAATCTTAAAGCGGTAGTTAGTTAATGCAATGCAACTATTTATGTTTACCTCAAGCC 425
Oy 358 TGCTCCCAAGCGGATTAAGCTCCGTACAAAGAGGAAATCAATCGACAAAGCAAGAGCA 417
Db 426 AGGTAGTAAGCGCAAAACATTGCAACCAACAAATTTGCTGAGCAAGTAGCCAAAG 485
Oy 418 TAGTCAACATCGTGAAG-----GTGGAAGCTCCAAAGAACGA 453
Db 486 AACTAAAGAGCTTAAGAAAAAGTTAGCTCAAGTGCCCATCTCAGTAAAGAAAGAGT 545
Oy 454 TGGTCTGTTGCTTGGCAGCGTTCCGCAAGAGAGCTTATCTACAGATGATGTTATATCTT 513
Db 546 TCGGGCAGTCAATGAAGCAAAAAAGACAGGAGCTATCTACAGACGATGGCTATATTTT 605
Oy 514 TAATGCTTCTGATATCATAGAGATACGTGATGCTTATATCGTTCTCATGAGATCA 573
Db 606 TAGTCCGACAGATATCATTTGATGATTTAGAGAGACGCTTATTTAGTACCTCATGTAATCA 665
Oy 574 TTACCATTAACATTCCTTAAGATGAGTTATCAGCTAGCGAGTTGGCTGCGAAGCCCTT 633
Db 666 CTATCATTATATTTCTTAAAAAGATTTGTCTCAGGTGAGCTAGCTGCTGCAAGCTTA 725
Oy 634 C 634
Db 726 C 726
```

Search completed: November 13, 2003, 21:21:13
Job time : 726 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2003, 20:59:48 ; Search time 4993 Seconds
(without alignments)
11628.959 Million cell updates/sec

Title: US-09-765-271-55

Perfect score: 2389
Sequence: 1 TTCTTACGAGTTGGACTGT.....TAAGTAGGAAAAATAAAC 2389

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%
Maximum Match 100%

Listing first 100 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estrov:*
6: em_estrpl:*
7: em_estro:*
8: em_hc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estrom:*
17: em_gse_hum:*
18: em_gse_hiv:*
19: em_gse_pin:*
20: em_gse_vrt:*
21: em_gse_fun:*
22: em_gse_mam:*
23: em_gse_mus:*
24: em_gse_pro:*
25: em_gse_rtd:*
26: em_gse_phg:*
27: em_gse_vrt:*
28: gb_gse1:*
29: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	70.8	3.0	906	28	BH153606 ENTTS83TF
2	69.8	2.9	890	28	BH146886 ENTTPK48TF
3	67.8	2.8	180	28	AF325819 AF325819
4	67.8	2.8	908	28	AZ548467 AZ548467 ENTERK30TR

5	66.4	2.8	829	29	BX173672
6	64.8	2.7	843	28	AZ551618
7	63.8	2.7	877	28	AZ531291
8	62.2	2.6	912	28	AZ551092
9	60.8	2.5	849	28	AZ546009
10	59.2	2.5	931	28	BH160272
11	58.6	2.5	1101	29	CNS00390
12	57.2	2.4	900	28	AZ549980
13	56.8	2.4	905	28	AZ550256
14	56.6	2.4	645	12	BH165350
15	56.6	2.4	747	12	BH162732
16	55.2	2.3	816	28	AZ535744
17	55	2.3	647	12	BH163120
18	55	2.3	671	12	BH168252
19	55	2.3	694	12	BH168242
20	55	2.3	717	12	BH160500
21	55	2.3	756	12	BH162492
22	55	2.3	794	12	BH159906
23	55	2.3	795	28	AZ528485
24	55	2.3	823	28	AZ676218
25	55	2.3	843	28	BH139532
26	53.8	2.3	721	13	BH141179
27	53.6	2.2	656	13	BH144040
28	53.6	2.2	844	29	BX139987
29	53.2	2.2	752	14	CD099026
30	52.8	2.2	671	13	BM273704
31	52.8	2.2	890	28	AZ530768
32	52.4	2.2	677	12	BH161314
33	52.4	2.2	725	12	BH169486
34	52.4	2.2	866	28	AZ527885
35	52.4	2.2	976	28	BH149983
36	52.2	2.2	945	29	CNS008CA
37	51.8	2.2	435	12	BM169710
38	51.8	2.2	880	28	AZ529191
39	51.6	2.2	853	29	CNS023KH
40	51	2.1	500	9	AU088119
41	50.8	2.1	726	28	AZ196050
42	50.2	2.1	1007	29	CNS06X9S
43	50	2.1	445	12	BM167469
44	50	2.1	555	13	BH496963
45	49.8	2.1	707	12	BM164483
46	49.6	2.1	598	12	BM170666
47	49.6	2.1	605	12	BM163520
48	49.4	2.1	629	12	BH073162
49	49.4	2.1	687	12	BM169665
50	49	2.1	1044	13	BX415231
51	48.8	2.0	1101	29	CNS0022U
52	48.6	2.0	605	12	BM162501
53	48.6	2.0	861	13	BM266138
54	48.4	2.0	546	13	BH497202
55	48.4	2.0	891	28	AZ683582
56	48.4	2.0	960	29	CC143364
57	48.2	2.0	468	10	BG553005
58	48.2	2.0	537	13	BH498263
59	48.2	2.0	554	12	BH275335
60	48.2	2.0	648	10	BE342803
61	48.2	2.0	661	9	AU060703
62	48.2	2.0	679	12	BH440521
63	48.2	2.0	756	12	BM165937
64	48.2	2.0	782	13	BM117562
65	48	2.0	548	12	BM159818
66	48	2.0	1101	29	CNS00L72
67	47.8	2.0	713	13	BH41428
68	47.8	2.0	718	13	BH257897
69	47.8	2.0	773	13	BH249618
70	47.8	2.0	792	13	BH14945
71	47.8	2.0	799	13	BH332339
72	47.8	2.0	834	12	BH090507
73	47.8	2.0	884	13	BH137322
74	47.6	2.0	840	28	AZ669474
75	47.4	2.0	541	29	CNS04KUS
76	47.4	2.0	731	13	BM115605
77	47.2	2.0	311	14	CA6999402

```

78 47.2 2.0 593 13 B0596786 PFESTOab2
79 47.2 2.0 828 13 BM249956 BM249956
80 47.2 2.0 871 13 BM262815 BM262815
81 47 2.0 530 12 BM165497 BM165497
82 47 2.0 838 28 A2535789 A2535789
83 46.8 2.0 740 9 AV679268 AV679268
84 46.8 2.0 1201 13 BX366070 BX366070
85 46.6 2.0 825 12 BX374097 BX374097
86 46.6 2.0 889 28 BH132245 BH132245
87 46.6 2.0 890 28 BH096477 BH096477
88 46.6 2.0 935 29 CNS00K01 CNS00K01
89 46.6 2.0 997 29 CNS005TE CNS005TE
90 46.4 1.9 507 28 A2167278 A2167278
91 46.4 1.9 692 13 BM130827 BM130827
92 46.4 1.9 761 13 BM120665 BM120665
93 46.4 1.9 947 28 BH148703 BH148703
94 46.2 1.9 838 13 BU129958 BU129958
95 46.2 1.9 861 13 BU129493 BU129493
96 46.2 1.9 888 28 A2528430 A2528430
97 46 1.9 426 13 BU495802 BU495802
98 46 1.9 546 28 A0008275 A0008275
99 46 1.9 714 12 BM170199 BM170199
100 46 1.9 916 28 A2671886 A2671886

```

ALIGNMENTS

```

RESULT 1          906 bp  DNA  linear  GSS 24-SEP-2001
LOCUS      BH153606/c
DEFINITION  ENTTS83TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            BH153606
            genomic survey sequence.
ACCESSION  BH153606.1 GI:15725323
VERSION    GSS.
SOURCE     Entamoeba histolytica
            Entamoeba histolytica
ORGANISM   Eukaryota; Entamoebidae; Entamoeba.
            1 (bases 1 to 906)
REFERENCE   1 (bases 1 to 906)
            Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.
            Determination of clone end sequences from Entamoeba histolytica
            HMI1MSS sheared DNA library (2001)
            Unpublished
JOURNAL    Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: b.loftus@tigr.org
COMMENT     Clones are derived from the Entamoeba histolytica HMI1MSS sheared
            DNA library
            Seg primer: M3-Forward
            Class: shotgun
            High quality sequence start: 15
            High quality sequence stop: 733.
            Location/Qualifiers
                1..906
                    /organism="Entamoeba histolytica"
                    /mol_type="genomic DNA"
                    /strain="HMI1MSS"
                    /db_xref="taxon:5759"
                    /clone_lib="Entamoeba histolytica Sheared DNA"
                    /note="Vector: pHO51; Site 1: Bst I; Constructed at The
                    Institute for Genomic Research (TIGR), Rockville, MD.
                    Genomic DNA isolated from broth cultures of E. histolytica
                    using a method described by Clark and Diamond (Clark,
                    C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
                    method for isolate identification. Exp. Parasitol.
                    77:450.). The DNA was mechanically sheared to give a
                    tight size distribution (~2 kb). The v + i method used for
                    the library construction is described in detail in Smith,

```

H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

Query Match 3.0%; Score 70.8; DB 28; Length 906;
 Best Local Similarity 43.8%; Pred. No. 0.00011;
 Matches 309; Conservative 0; Mismatches 337; Indels 0; Gaps 0;

```

BASE COUNT      149 a      293 c      66 g      398 t
ORIGIN
1504 TGGCAAAACCAATTCGAATAGTACTGAAAGCAAGTTCGATTCCTCAATTAGC
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
791 TGATGAAGAAAGATGATGATGAAGAAAGATGATGAAGAAAGATGATGATGATG
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1564 TGATGAAGTAAACATGCTCAGTGGTTACATTTTGGATGAACATGATTAATCATG
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
731 TGATGAATGAAGATGATGATGATGAAGAAAGATGATGATGATGATGATGATG
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1624 AGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
671 AGAAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1584 TTCTGATGAAGAAAGTTGCAAGCTCAAGCTTATATCTAAAGAAAGATCTTACC
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
611 AGAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1744 ATCTCCAGACGCAATGTTAAAGCAATCCACCTGAGATGATGATGATGATGATG
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
551 TGATGAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1804 TCGGTGTAAGGGGAAAGCAATTCCTGCTTCGATTCATATATGTTGAGCATAC
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
491 TGATGAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1864 AGTTGAGGTTAAAGACGTTAATTTGATTTCTCATTAAGGATCATTAATTAAT
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
431 TGATGAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1924 ATTTGCTTTGTTTATGATGATGATGATGATGATGATGATGATGATGATGATG
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
371 TGATGAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATG
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1984 GTTTCGACGATTAAGTAACTAGTAAACACCTGACGACGTCACATTTAATGATG
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
311 AGATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATG
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2044 ATGGGGCAATGCCAGTGAAGCATGTGTTAAGGCAAGAAACCAAGTAAAGTAA
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
251 AGACGACGATGAAGACGATGAAGACGATGAAGACGATGAAGACGATGAAGAC
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2104 GAACCTCAAGCGATGAAGAGCCAGTGAAGAAACACCTGCTAGCCAGAAATCC
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
191 AGAAGACGATGAAGAAAGCGATGAAGAAAGCGATGAAGAAAGCGATGAAGAA
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2164 AGTAGAGCTGAAGAAAGTAGAAGCCCACTCAAGAGACGAAGTT 2209
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
131 AGAAGACGATGAAGAAAGCGATGAAGAAAGCGATGAAGAAAGCGAT 86
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 2          890 bp  DNA  linear  GSS 27-AUG-2001
LOCUS      BH146886
DEFINITION  ENTPK48TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            BH146886
            genomic survey sequence.
ACCESSION  BH146886.1 GI:15302963
VERSION    GSS.
SOURCE     Entamoeba histolytica
            Entamoeba histolytica
ORGANISM   Eukaryota; Entamoebidae; Entamoeba.
            1 (bases 1 to 890)
REFERENCE   1 (bases 1 to 890)
            Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.

```

TITLE
Determination of clone end sequences from Entamoeba histolytica
JOURNAL
HM1:IMSS sheared DNA library (2001)
COMMENT
Unpublished
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@igf.org

Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
DNA library
Seq primer: ML3-Forward
Class: shotgun
High quality sequence strat: 17
High quality sequence scop: 880.
Location/Qualifiers

FEATURES

1..890
/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: PHOS1; Site 1: Set I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaubin and B.
Bartell, Oxford University Press, 1999)."

BASE COUNT 404 a 65 c 285 g 136 t
ORIGIN

Query Match 2.9%; Score 69.8; DB 28; Length 890;
Best Local Similarity 43.6%; Pred. No. 0.00018;
Matches 311; Conservative 0; Mismatches 402; Indels 0; Gaps 0;

1497 AGCGACTTGGCAACCAATCTCAATGATATAGTGAAGCAAGTTGCTATGCTC 1556
Db 15 AGAGATGATGATGAAGAGATGATGAAGAGATGATGAAGAGATGATGAAG 74
Qy 1557 AATTAGCTGATTAATACAGTCAGATGTTTATGATGAAGCATGATATATCA 1616
Db 75 ATGACGATGATGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 134
Qy 1617 GTGATGAAGGATGATGATGTAAGCCTCATATGCGCATATGTCATGAGTGAAG 1676
Db 135 AGCATGAAGAAACATGTAAGAAAGCATGAAGAAAGCATGAAGAAAGCATGAAG 194
Qy 1677 ATAGCCTTCTGATTAAGAAAAGTTGACGCTCAAGCCATATCTAAGAAAAGTATCC 1736
Db 135 AGCATGAAGAAACATGTAAGAAAGCATGAAGAAAGCATGATGATTTGAATTG 254
Qy 1737 TACCTCATCTCCAGACGAGATGTTAAAGCAATCCATCGAGTATGACAGACTA 1796
Db 255 AAGAGATGATGAAGAGAGATGATGAAGAAAGCATGATGAAGAGAGATGAAG 314
Qy 1797 TTTAATATGTTGTAAGAGGGAAGAAAGCAATTCCTGCTGATCTCATATATGTTG 1856
Db 315 AAGAGATGATGAAGAGAGATGATGAAGAAAGCATGATGAAGAGAGATGAAG 374
Qy 1857 AGCATAGATGATGAAGAGAGATGATGAAGAGATGATGAAGAGATGAAGAG 1916
Db 375 AAGAGATGATGAAGAGAGATGATGAAGAGATGATGAAGAGATGAAGAG 434
Qy 1917 ATATTAAATTTGCTGTTGATGATCAACATACAAAGCTCAATGCTTATACCTTGG 1976

Db 435 AAGAGATGATGAAGAGAGATGATGAAGAGAGATGATGAAGAGAGATGAAG 494
Qy 1977 AAGATTTGTTGGCCAGCATTAAGTACTAGTGAACACCTGAGCAAGTCCATTTCA 2036
Db 495 AAGATGAAGATGATGAAGAGATGATGAAGAGAGATGATGAAGAGATGAAGAG 554
Qy 2037 ATGATGATGAGGCGCAATCCAGTGAAGATGTTTATGAGCAAGAAAGCAAGTGAATC 2096
Db 555 ACGATGAAGAGAGATGAAGAGAGATGAAGAGAGATGAAGAGAGATGAAGAG 614
Qy 2097 CAATTAAGACTTCAAGCGATGAAGAGAGATGAAGAGAGATGAAGAGAGATGAAG 2156
Db 615 ACGATGAAGAGAGATGAAGAGAGATGAAGAGAGATGAAGAGAGATGAAGAG 674
Qy 2157 TCCCTCAAGTGAAGCTGAAGAAAGTGAAGAGAGATGAAGAGAGATGAAGAG 727
Db 675 ACGATGAAGAGAGATGAAGAGAGATGAAGAGAGATGAAGAGAGATGAAGAG 727

RESULT 3
AF325819
LOCUS
DEFINITION
AF325819 Leptospira interrogans serovar lai DNA Leptospira
interrogans serovar lai genomic clone 18, genomic survey sequence.
ACCESSION
AF325819
VERSION
AF325819.1 GI:19263459
KEYWORDS
GSS.
SOURCE
ORGANISM
Leptospira interrogans serovar lai
Leptospira interrogans serovar lai
Bacteria; Spirochaetes; Spirochaetales; Leptospiraceae; Leptospira.
REFERENCE
1 (bases 1 to 180)
Hu, C. and Bao, L.
Leptospira interrogans serovar lai DNA
UNPUBLISHED
CONTACT: Hu C
Leptospirosis Research Unit
West China University of Medical Sciences
17# Renmin Nan Road, Chengdu, Sichuan 610041, P.R. China
Email: huchanghua@263.net
CLASS: unknown.

FEATURES
source
Location/Qualifiers

1..180
/organism="Leptospira interrogans serovar lai"
/mol_type="genomic DNA"
/db_xref="taxon:57678"
/clone="18"
/clone_lib="Leptospira interrogans serovar lai DNA"
/note="Obtained through DNA subtraction using the
nonpathogenic strain Leptospira biflexa serovar Patoc
strain Patoc I"

BASE COUNT 66 a 27 c 40 g 47 t
ORIGIN

Query Match 2.8%; Score 67.8; DB 28; Length 180;
Best Local Similarity 62.9%; Pred. No. 0.00029;
Matches 105; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

249 TACTCATTAAGATCAATTAAGATTAAGATTAAGATTAAGATTAAGATTAAGATTAAG 308
Db 1 TATTATATGACGACCCCAATTAATCAATTAAGATTAAGATTAAGATTAAGATTAAG 60
Qy 309 GTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 368
Db 61 ATGGCTATATTAATCAAGTTGATGCAAGTATTAATGTTATTAATTAAGCCAGAAAGC 120
Qy 369 CGATTAAGCTCCGTACAAAGAGAAATCAATCAAGCAAGCAAGCAAGCAAGCAAG 415
Db 121 GTAAATATATCGTATCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 167

RESULT 4
A2548467
LOCUS
A2548467 908 bp DNA linear GSS 14-NOV-2000

DEFINITION ENTEK30TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, genomic survey sequence.

ACCESSION A2548467

VERSION A2548467.1 GI:11172102

KEYWORDS GSS.

SOURCE Entamoeba histolytica

ORGANISM Entamoeba histolytica

REFERENCE 1 (bases 1 to 908)
Loftus, B., Van Aken, S. and Fraser, C.
Determination of clone end sequences from Entamoeba histolytica
Unpublished

JOURNAL
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@igr.org
Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library
Seq primer: M13-Reverse
Class: Shotgun
High quality sequence start: 17
High quality sequence stop: 828.
Location/Qualifiers

FEATURES
source
1..908
/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HM1:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHOSt1; Site_1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 434 a 16 c 289 g 169 t

ORIGIN

Query Match 2.8%; Score 67.8; DB 28; Length 908;
Best Local Similarity 44.0%; Pred. No. 0.00049;
Matches 334; Conservative 0; Mismatches 422; Indels 3; Gaps 1;

1525 TGAGTATCTGGAAGACGAAGTTCGATTCCTCAATTAGCTGATTAAGTATCAACGCTCAGA 1584
33 TGAAGATGATGTGATGAAGAGATGATGATGAAGAGATGTGATGAAGAGATGATGA 92

1555 TGGTTACATTTTGTGTAACATGATATATCATGATGAAGAGATGCATATGTACGCC 1644
93 TGATGAAGAAGATGATGAAGAAGATGATGAAGAAGATGATGAAGAAGATGA 152

1645 TCATATGGGCGCTACTGCTGATG---GAAAGATAGCCTTCTGATTAAGGAAAAAGT 1701
153 TGAAGACGATGAAGACGACGAATTAAGATTAAGATGATGATGAAGAAGAAAGA 212

1702 TGCACTCAAGCCTTACTTAAGAAAAAGTATCTTCACTCCATCTCCAGCGCAGATGT 1761
213 TGATGATGAAGATGATGAAGAGATGAAGACGAATTAAGATTAAGAGATGATGA 272

1762 TAAAGCAATTCACATGAGATAGTGCACGACCTATTTCATTCCTGTGAAAGGGGAAA 1821
273 TGATGAAGAAGAGATGATGAAGAAGAAAGATGATGATGAAGAAGATGATGAAGA 332

1822 ACGAATTCACCTGCTTCGACTTCATATATGTTGACCATACAGTTGAGTTAAAAACGG 1881
333 AGAAGATGAAGAAGATGATGAAGAAGATGATGAAGAAGATGATGAAGAAGATGA 392

1882 TAATTTGATTTATTCCTCATAGGATCTTACCATTAATTAATTTCTTGTTGATGA 1941
393 TGATGAAGATGATGAAGAAGATGATGAAGAAGATGATGAAGAAGATGATGAAGA 452

1942 TCACATTAACAAGCTCCAAATGCGCTTACCTTGGAAGATTTGTTGCGACGATTAA 2001
453 AGAAGAAGATGATATGAAGAAGATGATGAAGAAGATGATGAAGAAGATGA 512

2002 CTACGTGAACACCCCTGACGACGCTCCATTTCAATGATGATGAGGCGCAATGCCAGTGA 2061
513 TGAAGATGATGATGAAGAAGATGATGAAGAAGATGATGAAGAAGATGATGAAGA 572

2062 GCATGTGTTAGGCAAGAAGACCACTGTAAGATCCAAATTAAGAACTTCAAGCGGATGA 2121
573 AGAAGATGAAGAAGAAAGATGATGAAGAAGATGATGAAGAAGATGATGAAGA 632

2122 AGAGCCGTAGAGAAACACCTGCTGAGCCAGAGTCCCTCACTAGAGCTGAAAAGT 2181
633 AGACGACGAATGTAATTAAGATGAAGAAGATGATGAAGAAGATGATGAAGAAG 692

2182 AGAAGCCCAACTCAAGACGAGAGTTTGTCTGCGAAGTAAAGGATTTCTACTGAA 2241
693 AGATGATGAAGAAGAAAGATGATGATGATGATGAAGAAGATGATGAAGAAGATGA 752

2242 AGCCAATGCAACAGAACTCTAGCTGTTTACGAATPA 2280
753 TGAAGAAGATGAAGATGATGAAGATGATGAAGAAGATGA 791.

RESULT 5
BX173672/c 829 bp DNA linear GSS 13-MAR-2003

LOCUS BX173672
DEFINITION Dario rerio genomic clone DKEY-150M6, genomic survey sequence.
ACCESSION BX173672
VERSION BX173672.1 GI:28005377

KEYWORDS GSS.

SOURCE Dario rerio (zebrafish)

ORGANISM Dario rerio

REFERENCE 1 (bases 1 to 829)
Humphray, S.J., Huckle, E. and Durham, J.L.
Direct Submission
Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humphray@sanger.ac.uk Unpublished

COMMENT This sequence was generated from the S66 end of BAC 150M6. 150M6 is part of the Dariokey BAC library created by R. Plasterk and N.V. Keygene. Further details:
http://www.sanger.ac.uk/projects/D_rerio/.

FEATURES
source
1..829
/organism="Dario rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-150M6"
/issue="Testis"
/note="vector pIndigoBAC-536"

BASE COUNT 260 a 103 c 32 g 434 t

ORIGIN

Query Match 2.8%; Score 66.4; DB 29; Length 829;
Best Local Similarity 47.6%; Pred. No. 0.00097;
Matches 196; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

186 ATGCGACCACTATCATTAATTAAGTGAAGTTCCTTAAGCGCTATCATCAAGTAA 245
635 ATAAAGTAATTAATTAATGATGATGATGATGATGATGATTAATTAATTAATAGTA 576

QY 246 AATTACTCATGAAGATCCAACTATAGCTAAAGATGAGATATTGTTAAGAGTCA 305
 DB 575 ATATATATATATATATATAGAAATATATATATATATATATATATATATAT 516
 QY 306 AGGGTGGATATGTTATCAAGGATGATGAGAAATACTATGTTTACCTTAAGATGCTGCC 365
 DB 515 AT 456
 QY 366 AGCGGATACGTCCTACAAAAGAGATCAATGACAAAAGAGAGCATAGTCAAC 425
 DB 455 AT 396
 QY 426 ATCTGTAAGTGGAACTCCAGAAAGATGCTGCTGCTGCTGCTGCTGCTGCTG 485
 DB 395 AT 336
 QY 486 GCTATATCTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGTACTGTTG 545
 DB 335 GTAT 276
 QY 546 ATGCTATATGCTCTCATGAGATCATTTACATTAATCTTCTAAGATGA 597
 DB 275 AT 224
 RESULT 6
 AZ551618/c 843 bp DNA linear GSS 14-NOV-2000
 LOCUS ENT554TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
 DEFINITION genomic, genomic survey sequence.
 ACCESSION AZ551618
 VERSION AZ551618.1 GI:11176919
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica
 ORGANISM Entamoeba histolytica
 Eukaryota; Entamoebidae; Entamoeba.
 1 (bases 1 to 843)
 Loftus, B., Van Aken, S. and Fraser, C.
 Determination of clone end sequences from Entamoeba histolytica
 HMI:IMSS sheared DNA library
 Unpublished
 CONTACT: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@tigr.org
 Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
 DNA library
 Seq primer: M13-Reverse
 Clases: Shotgun
 High quality sequence start: 39
 High quality sequence stop: 838.
 Location/Qualifiers
 1. 843
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HMI:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica sheared DNA"
 /note="Vector: PHOS1, Site 1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.) The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + 1 method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Bartell, Oxford University Press, 1999."

BASE COUNT 155 a 266 c 18 g 404 t
 ORIGIN
 Query Match 2.7%; Score 64.8; DB 28; Length 843;
 Best Local Similarity 44.1%; Pred. No. 0.0022;
 Matches 319; Conservative 0; Mismatches 402; Indels 3; Gaps 1;

QY 1521 AATTGACTATATCTAAGACCAAGTTCCTATTTGCTCATTTGCTATTAAGTATACAGT 1580
 DB 753 AAGAAGAT 694
 QY 1581 CAGATGTTTACATTTTGGATGACATGATATATATATATATATATATATAT 1637
 DB 693 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 634
 QY 1638 TAAAGCTTATATGAGGACATATGTCATGATGATGATGATGATGATGATG 1697
 DB 633 ACGAATATGATATGATGATGATGATGATGATGATGATGATGATGATGATG 574
 QY 1698 AAGTTCAGCTTACACCTATATCTAAGAAAAAGATATCTTACTTCTCATCTC 1757
 DB 573 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 514
 QY 1758 ATGTTAAAGCAAAATCCACTGAGATAGTGCAGCAGTATTATTAATGCTGTA 1817
 DB 513 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 454
 QY 1818 AAAAAAGATTCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1877
 DB 453 AATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 394
 QY 1878 ACGTATATTTGATTTTCTCTATAGATCATTAATATATATATATATATAT 1937
 DB 393 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 334
 QY 1938 ATGATCACAATATCAAGAGTCCAAATGCTTACTTGGAGATTTGTTGGAGAT 1997
 DB 333 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 274
 QY 1998 AGTACTAGTGAACACCTGAGACCTGAGACCTGATTTATATGATGATGATGAT 2057
 DB 273 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 214
 QY 2058 GTGACATGTTGTTAGGCAAGAAAGCCACAGTGAAGATCCAAATTAAGATT 2117
 DB 213 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 154
 QY 2118 ATGAAGAGCCAGTGAAGAAACACCTGCTGAGCCGAGAGTCCCAATAGAG 2177
 DB 153 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 94
 QY 2178 AAGTGAAGCCCACTCAAGAGAGAGAGTTTCTGCGAAAGATTAACGATTTAG 2237
 DB 93 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 34
 QY 2238 TGAAT 2241
 DB 33 AGAA 30
 RESULT 7
 AZ531291/c 877 bp DNA linear GSS 03-NOV-2000
 LOCUS ENT5034TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
 DEFINITION genomic, genomic survey sequence.
 ACCESSION AZ531291
 VERSION AZ531291.1 GI:11085838
 KEYWORDS GSS.
 SOURCE Entamoeba histolytica
 ORGANISM Entamoeba histolytica
 Eukaryota; Entamoebidae; Entamoeba.

REFERENCE 1 (bases 1 to 877)
 Authors Loftus, B., Van Aken, S. and Fraser, C.
 Title Determination of clone end sequences from Entamoeba histolytica
 Journal HM:IMSS sheared DNA library
 Comment Unpublished
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@tigr.org
 Clones are derived from the Entamoeba histolytica HM:IMSS sheared DNA library
 Seq primer: M13-Reverse
 Class: shotgun
 High quality sequence start: 22
 High quality sequence stop: 829.
 Location/Qualifiers
 1..877
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HM:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: pHOSt1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. in Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 156 a 268 c 29 g 424 t
 ORIGIN

Query Match 2.7%; Score 63.8; DB 28; Length 877;
 Best Local Similarity 43.4%; Pred. No. 0.0036;
 Matches 293; Conservative 0; Mismatches 382; Indels 0; Gaps 0;

OY 1535 GAAGCGAAGTTGCTATTCCTCAATTAGCTGTAAGTATACACCTCAGATGCTTACATT 1594
 DB 785 GAAGAAAGATGATGATGATGATGAAGAAAGAAAGATGATGATGATGATGAAGAA 726
 OY 1595 TTTGATGAACATGATATATATCATGATGAAGAGATGATGATGATGATGATGATGATG 1654
 DB 725 GATGATGATGAAGAAAGATGATGATGAAGAAAGATGATGATGAAGAAAGATGATGAAGA 666
 OY 1655 CATAGTCACTGATGTAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATG 1714
 DB 665 GATGATGAAGAAAGAAAGATGATGATGAAGTATGATGAAGAGATGATGAAGACGACGAA 606
 OY 1715 TATATGAAGAAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1774
 DB 605 TATGATTAAGAAAGATGATGATGATGAAGAAAGAAAGATGATGAAGATGATGAAGAC 546
 OY 1775 ACTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1834
 DB 545 GATGAAGACGACGATGATGATGATGAAGATGATGATGATGATGAAGAAAGATGATGAT 486
 OY 1835 GTTCGACTTCATATATGTTGAGCATACAGTTGAGTTAAACGGTAAATTTGATTTATT 1894
 DB 485 GAAGAAAGAAAGAAAGATGATGATGAAGTATGATGAAGAAAGATGATGAAGATGATGA 426
 OY 1895 CCTATTAAGATCATTTACATATATTAATTTGCTGTTGATGATGATGATGATGATGATG 1954
 DB 425 GAAGATGATGAAGATGATGATGAAGAAAGAAAGATGATGATGAAGATGATGAAGACGAT 366

OY 1955 GCTCCAAATGGCTATACCTTGAGAGATTTGTCGACGATTAAGTACTACGTAAGACAC 2014
 DB 365 GAAGACGACGAAATGATGATTAAGAGATGATGATGAAGAAAGAAAGATGATGAAGAA 306
 OY 2015 CCTGACGACCTCCATCTTCAATGATGATGATGATGATGATGATGATGATGATGATG 2074
 DB 305 GAAGAAAGTATGATGATGAAGATGATGAAGAAAGAAAGATGATGATGAAGAAAGAA 246
 OY 2075 AAGAAAGCCACAGTGAAGATCCAAATTAAGACTTCAAGCGGATGAAGACCGTAGAG 2134
 DB 245 GATGAAGATGATGATGAAGAAAGATGATGATGAAGAAAGATGATGAAGAAAGAAAT 186
 OY 2135 GAACACCTGCTGACCGAAGATCCCTCAAGTACAGACTGAAGAAAGACCACTC 2194
 DB 185 GATGAAGAAAGATGATGAAGAAAGAAAGATGATGATGAAGACGACGAATATGAATTGA 126
 OY 2195 AAGAAAGCAAGATT 2209
 DB 125 GATGATGAAGAAAGAT 111

RESULT 8
 A2551092
 LOCUS
 DEFINITION ENTJ22TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
 accession A2551092
 version A2551092.1 GI:11176393
 keywords GSS.
 SOURCE
 ORGANISM
 Entamoeba histolytica
 Entamoeba histolytica
 Eukaryota; Entamoebidae; Entamoeba.
 REFERENCE
 1 (bases 1 to 912)
 Authors Loftus, B., Van Aken, S. and Fraser, C.
 Title Determination of clone end sequences from Entamoeba histolytica
 Journal HM:IMSS sheared DNA library
 Comment Unpublished
 Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@tigr.org
 Clones are derived from the Entamoeba histolytica HM:IMSS sheared DNA library
 Seq primer: M13-Forward
 Class: shotgun
 High quality sequence start: 17
 High quality sequence stop: 861.
 Location/Qualifiers
 1..912
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HM:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: pHOSt1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. in Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 423 a 34 c 264 g 191 t
 ORIGIN

Query Match	2.6%;	Score 62.2;	DB 28;	Length 912;
Best Local Similarity	43.3%;	Pred. No. 0.0082;		
Matches 292;	Conservative	0;	Mismatches 383;	Indels 0;
				Gaps 0;

Tel: 301 838 0208
Fax: 301 838 3543
Email: bjloftus@tigr.org

Clones are derived from the *Entamoeba histolytica* HM1:IMSS sheared DNA library
Seq primer: M13-Forward
Class: shotgun

FEATURES
source

Location/Qualifiers
1. .849

```

/organism="Entamoeba histolytica"
/mol_type="genomic DNA"

```

```

/sciadin="HMI:IMSS"
/db xref="taxon:5759"
/citation lib="Entamoeba histolytica shared DNA"

```

BASE COUNT
ORIGIN

199 a	236 c	43 g	371 e
-------	-------	------	-------

Query Match	2.5%;	Score 60.8;	DB 28;	Length 849;
-------------	-------	-------------	--------	-------------

Qy	1190	TATGCTCTTTGGGAAAGATTTTAAACCATCTGAAACCTGTAAAAATCTTTGAAAGCAAGCTTTATCA	1249
Db	742	TTTGACTTTAGATGAAAAATGTGAAATGGAACATGAATACCAGATGATGAAAGCAATGAATTA	683
Qy	1250	AAACCAAGAGAGTGTTCACACACTTTTAACGTCTAAAAAAGAAAAATGTTGTCTCTCGTGAC	1309
Db	682	GAAAGAAATGATGATGTGATGAATTTGAATTGAAAGCAAGAAAGATGTGATGATGATGAA	623
Qy	1310	CAAGAAATTTTATGATTAAGCATATAATCTGTAACTGAGGCTCATMAAGCTTGTTTGMA	1366
Db	622	GACCAAGAAAGATGAAAGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	563
Qy	1370	AATAAAGGTGTAATTCGATTTCCAGACCTTAGACAAATTTATAGAACGTTGAATGAT	1422
Db	562	GAAATTTGAATTTAGAGATGAAGAAAGATGAGATGATGATGATGATGATGATGATGATGAT	503
Qy	1430	GAATCGACTAATAAGAAAAAATTTGATGATGATTTATTTGCGATTCCTAGCAACAATTACC	1488
Db	502	GAAATGATGATGATGAAAGATGAAAGAAAGATGATGATGATGAAAGAACGAAGAAAGATGAT	443
Qy	1490	CATCCAGAGCGACTTGGCAAAACCAATTTCCAAATTGAGTAACTGAAGACGAAGTTGCT	1549
Db	442	GATGATGAATTTGATTAAGAAAGATGAAAGAAAGATGAAAGAAAGATGATGATGAT	383
Qy	1550	ATTGCTCAATTAGCTGATAGATATACAAACGTCAAGATGTTACATTTTGTGATGAACATGAT	1606
Db	382	GATGATGAAAGATGATGAAGATGATGAAGATGATGATGATGATGATGATGATGATGATGAT	323
Qy	1610	ATAATCAAGTGTGAAGGAGATGCAATGTGAACGCTCATATGGGCGAATGCACATGATTT	1665
Db	322	GAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	263
Qy	1670	GGAAGAAATAGCTTTCTGATTAAGAAAAAATTTGAGCTCAAGCTCAAGCTTAACTAAAGAAAA	1722
Db	262	GAAATGAAAGAAATGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	203
Qy	1730	GGTATCCTACCTCATCTCCAGACGCAAGATGTTAAAGCAAAATCCAACTGGAGATGTGCA	1788

Db	202	GCATTGGAAGATGGAATGACATGATGAAGTGAAGAAAGATGAAGATGATGAAGAA	143
Oy	1790	GCAGCTATTTTACAATCGTGTGAAAGCGGAAAA	1822
Db	142	GAAAGAAAGATGATGATGATGAAGTGAAGAA	110
RESULT 10	BH160272	931 bp	linear
LOCUS	BH160272		GSS 24-SEP-2001
DEFINITION	ENTOV49TR Entamoeba histolytica Sheared DNA	Entamoeba histolytica	
ACCESSION	BH160272		
VERSION	BH160272.1	GI:15733710	
KEYWORDS	GSS.		
ORGANISM	Entamoeba histolytica		
REFERENCE	Entamoeba histolytica		
AUTHORS	1 (bases 1 to 931)		
TITLE	Loftus, B., Wang, G., Van Aken, S., and Fraser, C.		
JOURNAL	Determination of clone end sequences from Entamoeba histolytica		
COMMENT	HM1:IMSS sheared DNA library (2001)		
FEATURES	Unpublished		
SOURCE	Contact: Brendan J Loftus		
	Department of Eukaryotic Genomics		
	The Institute for Genomic Research		
	9712 Medical Center Dr., Rockville, MD 20850, USA		
	Tel: 301 838 0208		
	Fax: 301 838 3543		
	Email: b.loftus@igrr.org		
	Clones are derived from the Entamoeba histolytica HM:IMSS sheared		
	DNA library		
	Seq primer: M13-Reverse		
	Class: shotgun		
	High quality sequence start: 8		
	High quality sequence stop: 677.		
	Location/Qualifiers		
	1..931		
	/organism="Entamoeba histolytica"		
	/mol_type="genomic DNA"		
	/strain="HM1:IMSS"		
	/db_xref="taxon:5759"		
	/clone_lib="Entamoeba histolytica Sheared DNA"		
	/note="Vector: pHOSt1; Site 1: Bst I; Constructed at The		
	Institute for Genomic Research (TIGR), Rockville, MD.		
	Genomic DNA isolated from broth cultures of E. histolytica		
	using a method described by Clark and Diamond (Clark,		
	C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a		
	method for isolate identification. Exp. Parasitol.		
	77:450.). The DNA was mechanically sheared to give a		
	tight size distribution (~2 kb). The v + i method used for		
	the library construction is described in detail in Smith,		
	H.O. and Venter, J.C. (Making small insert libraries for		
	whole genome shotgun sequencing projects. in Genome		
	Sequencing: A Practical Approach, eds. M. Vaundin and B.		
	Barelli, Oxford University Press, 1999)."		
BASE COUNT	402 a	279 g	199 t
ORIGIN	51 c		
Query Match	2.5%	Score 59.2;	DB 28; Length 931;
Best Local Similarity	43.3%	Pred. No. 0.037;	
Matches 277; Conservative	0;	Mismatches 363;	Indels 0;
		Gaps 0;	
Oy	1583	GATGGTTCATTTTGGATGACATGATATTAATCACTGATGGAAGGATGCATATGTACG	1642
Db	17	GATGATGATGGAAGAAAGATGATGATGATGAAGTGAAGACGATGAAGACACGAA	76
Oy	1643	CCTCATATGGGCGCATAGTCATCGATTTGGAAAAAGATACCTTTCTGATPAGAAAAAGTT	1702
Db	77	TATGATATGAAAGATGATGATGATGATGAAGAAAGATGATGATGAAGAAAGAAAT	136
Oy	1703	GCAGCTCAAGCCTTACTATGAAGAAAAAGATATCTCACTCCATCTCCACAGCGAGATGTT	1762

Db	137	GATGATGAAAGTGTATGATGAAGAAAGACATGAAGAAGTGTGAAGAAGATGATGAAGT	196
Oy	1763	AAAGCAATCCCACTGAGATAGTGCAGCAGCTATTTAACATCGTGTGAAAGGGGAAAA	1822
Db	197	GATGATGAAGAAGAAGATGATGATGAAGATGATGAAGACGATGAAGACGACGAATAT	256
Oy	1823	CGAATTCCACTCGTTCCACTTCCATATATGTTGAGCATACAGTTGAGCTTTAAAAACGT	1882
Db	257	GAAATTGAGAAGTATGATGAAGAAGAAGAACATATGATGAAGAAGAAGATGATAT	316
Oy	1883	AATTGATTTATCTCTCAATAGATCATTCATTAATTAATTGCTTGGTTTGATGAT	1942
Db	317	GAAATGATGATGATGAAGAAGATGATGATGATGAAGAAGATGATGAAGATGATAT	376
Oy	1943	CACACATACAAAGCTCCAAATGGCTATACCTTGAAGATTTGTTGCGACGATTAAGTAC	2002
Db	377	GAAAGAGAAGTATGATGAAGAAGAAGATGAAGAAGAAGATGATGAAGAAGATGAT	436
Oy	2003	TACGTAGAACACCTCGACGACGCTCACATTTCTAATGATGATGGCGAATCCAGTAG	2062
Db	437	GAAAGAGAAGTATGATGAAGAAGATGATGAAGAAGATGATGAAGAAGATGATGAAGAAGAT	496
Oy	2063	CATGTGTTAGCAAGAAGACACAGTAGATGCCAATTAAGAACTTCGAACGGATGAA	2122
Db	497	GATGAAGAAGTATGATGAAGAAGATGATGAAGAAGATGATGAAGAAGATGATGAT	556
Oy	2123	GAGCCAGTAGAGAAACACCTGCTGAGCAGAAAGTCCCTCAAGTAGAGACTGAAAAAGTA	2182
Db	557	GAAAGAAGTATGATGAAGAAGATGATGAAGAAGATGAAGATGATGAAGAAGATGATAT	616
Oy	2183	GAAGCCCACTCAAGAAGACGAAGTTTGTCTTGCGAAG	2222
Db	617	GAAAGATGATGAAGTATGATGAAGAAGATGATGAAGATGATGAAG	656

RESULT 11

CNS00396/c

LOCUS

DEFINITION

Drosophila melanogaster genome survey sequence TE13 end of BAC # BACR08K10 of RPC1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

AL063921

AL063921.1 GI:4941778

GSS.

Drosophila melanogaster (fruit fly)

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephyndroidea; Drosophilidae; Drosophila.

1 (bases 1 to 1101)

Genoscope.

Direct Submission

Submitted (02-JUN-1999) Genoscope - Centre National de Sequençage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr

- Web : www.genoscope.cns.fr)

Determination of this BAC end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mammeter in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPC1-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

SOURCE

1..1101

/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone_lib="BACR08K10"
/note="end : RPCT-98"
/note="end : TE3"

BASE COUNT 201 a 64 c 131 g 202 t 503 others

ORIGIN

Query Match 2.5%; Score 58.6; DB 29; Length 1101;
Best Local Similarity 16.2%; Pred. No. 0.053;
Matches 97; Conservative 252; Mismatches 251; Indels 0; Gaps 0;

1355 AAGCCTTGTGNAATAGGCTGTAATTCGATTTCCAGCCTTAGACAATATTA 1414
1098 RRMGGDTWRDTRKDDMDWTMMWKDADRRMAGDADRWAMWDGATWTTATMWW 1039
1415 GAACCTGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1474
1038 WMMATWTDWMDKMMWMTAAKTDTATWMTAMRADWAGRDGAKRDADATDAGAR 979
1475 CTAGACCAATTAACCATCCAGAGCAGCTTGCAACCAATTCGAATTCAGTACT 1534
978 RGGRRKK 919
1535 GAGAGCAAGTTCGATTCCTCATTCAGTTCAGTTCAGTTCAGTTCAGTTCAGT 1594
918 DGGAGDDKDDGKGDADDDTGTGKDDDDKDDKDDKDDKDDKDDKDDKDDKDD 859
1595 TTTGATGAACATGATATTAATCAGTGATGAAGAGATGATGATGATGATGATG 1654
858 DADWMTWDAADDDWMDADDDWMDADDDWMDADDDWMDADDDWMDADDDWMD 799
1655 CATAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1714
798 RADDDKDDADDDDDAATTTTTRTDDTDDWMDKKTDTWRMAADTTWRDDDD 739
1715 TTTACTAAGAAAAGATATCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1774
738 TGRKKRRRTWRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRRKRR 679
1775 ACTGAGATAGTGCAGACCTATTTTCAATCGTGAAAGGGGAAAAAGAAATTC 1834
678 TWRDADWADADWMTTDTDTDDMDKRRRRKRRRRRTTARAAWMDWTKAMWMA 619
1835 GTTCACTTCATATATGTTGAGCTACATGATGATGATGATGATGATGATGAT 1894
618 DKTRADRWDAADWTDARADRWDAADWTDARADRWDAADWTDARADRWDAAD 559
1895 CCTCAATGAGATCATACATATATTAATTTGCTGTTGATGATGATGATGATG 1954
558 TTAARAAAWMAWMAWMTTATTTTATTTTATTTTATTTTATTTTATTTTAT 499

RESULT 12 900 bp DNA linear GSS 14-NOV-2000

AZ549980 Entamoeba histolytica Sheared DNA Entamoeba histolytica

DEFINITION AZ549980 genomic survey sequence.

ACCESSION AZ549980

VERSION AZ549980.1 GI:11175122

KEYWORDS GSS.

SOURCE Entamoeba histolytica

ORGANISM Entamoeba histolytica

REFERENCE 1 (bases 1 to 900)

AUTHORS Loftus, B., Van Aken, S., and Fraser, C.

TITLE Determination of clone end sequences from Entamoeba histolytica

JOURNAL HMI:IMS sheared DNA library

COMMENT Unpublished

Contact: Brendan J Loftus

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@igr.org
Clones are derived from the Entamoeba histolytica HMI:IMS sheared
DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 20
High quality sequence stop: 890.
Location/Qualifiers

FEATURES source

1. 900
/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HMI:IMS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHS1; Site 1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O., and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaubin and B.
Bartell, Oxford University Press, 1999)."

BASE COUNT 376 a 63 c 235 g 226 t

ORIGIN

Query Match 2.4%; Score 57.2; DB 28; Length 900;
Best Local Similarity 45.5%; Pred. No. 0.1;
Matches 203; Conservative 0; Mismatches 243; Indels 0; Gaps 0;

166 ATGGGACCATCATATATTATTAATGATGATGATGATGATGATGATGATGATG 245
380 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 439
246 AATTACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 305
440 ATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 499
306 ACGGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 365
500 ATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 559
366 ACCGGATTAACGTCCTGTCACAAAAGGAAATCAATCAACAAAAGAGCATAT 425
560 AAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATG 619
426 ATCGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 485
620 ATGAAGATGAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATG 679
486 GCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 545
680 ACGATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 739
546 ATGCTATATGTTTCTGATGATGATGATGATGATGATGATGATGATGATGAT 605
740 ACGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATGATGAAGATG 799
606 CTAGCGAGTGGCTGCTGAGAGCC 631
800 ACGATGATGAAGATGATGAAGC 825

RESULT 13 905 bp DNA linear GSS 14-NOV-2000

AZ550256/c

LOCUS

DEFINITION ENTEV58TR Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, genomic survey sequence.

ACCESSION AZ550256

VERSION AZ550256.1 GI:11175557

KEYWORDS GSS.

SOURCE Entamoeba histolytica

ORGANISM Entamoeba histolytica

REFERENCE 1 (bases 1 to 905)

AUTHORS Loftus, B., Van Aken, S. and Fraser, C.

TITLE Determination of clone end sequences from Entamoeba histolytica HMI:IMSS sheared DNA library

JOURNAL Unpublished

COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library

Seq primer: M13-Reverse

Class: Shotgun

High quality sequence start: 100

High quality sequence stop: 872.

Location/Qualifiers

1..905

/organism="Entamoeba histolytica"

/mol_type="genomic DNA"

/strain="HMI:IMSS"

/db_xref="taxon:5759"

/clone_lib="Entamoeba histolytica Sheared DNA"

/note="Vector: PHOSI, Site1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 203 a 260 c 46 g 396 t

ORIGIN

Query Match 2.4% Score 56.8; DB 26; Length 905;

Best Local Similarity 44.8%; Pred. No. 0.12;

Matches 269; Conservative 0; Mismatches 323; Indels 9; Gaps 1;

1139 GTACGAAAAGTTGGGAGAGATATGTATTCGAGAAAAGGCGATCTCTGTTATGCTTT 1198

DB GAATTAGGAAGAGATGATGATGATGAATTGGAATTAGGAAGAGATGATGATGAT 721

1139 GCGAAAGATTACCATCTGAACCTTTAAATCTTGAAGCAAGTATCAAAACAAGG 1258

DB GATGGAAGCAGGAAGATGAGAAAGAGATGATGATGATGATGATGATGATGAT 661

1259 AGTGTTCACACACTTAACTGTAAGAAAAGAAATGTTGCTCTCGTACCAAGATT 1318

DB GATGATGAATTTGAATTGAGATGAGAGATGAGATGATGATGATGATGATGAT 601

1319 TATGATTAAGCATATATCTGTTAACTGAGGCTCATAT-----AAGCTTGTGTTA 1369

DB GATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 541

1370 AATTAAGGTCGTAATTCGATTTCCAGGCTTGAACAAATTAATTGAACGCTTGATGAT 1429

DB GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 481

1430 GAATCGACTATATAAGAAAATTGGTAGATGATTTATGGCATTCCTAGCAACCAATTACC 1489

DB GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 421

1490 CATCCAGAGCAGCTTTGGCAACCAATTCGAATTGATGATGATGATGATGATGAT 1549

DB GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 361

1550 ATTGCTCAATTAAGCTGTTAAGTATATACACGTCAGATGTTTCATTTTGTATGATGAT 1609

DB GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 301

1610 ATAACTAGTATGAGAGAGATGATGATGATGATGATGATGATGATGATGATGAT 1669

DB TTAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 241

1670 GGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1729

DB GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 181

1730 G 1730

DB 180 G 180

RESULT 14

LOCUS BML65350 645 bp mRNA linear EST 04-DEC-2001

DEFINITION EST567873 PYBS Plasmodium yoelii yoelii cDNA clone PYCML60 5' end, mRNA sequence.

ACCESSION BML65350

VERSION BML65350

KEYWORDS BML65350.1 GI:17311031

SOURCE EST.

ORGANISM Plasmodium yoelii yoelii

REFERENCE Plasmodium yoelii yoelii

AUTHORS Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium. 1 (bases 1 to 645)

Fraser, C.M., Daly, T.M., Long, C.A., Bergman, L.W., Vaidya, A.B., Plasmodium yoelii EST project at TIGR

TITLE Unpublished

JOURNAL Contact: Jane Carlton

COMMENT Parasite Genomics Group

The Institute for Genomic Research

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-530-9319

Fax: 301-838-0208

Email: carlton@tigr.org

For clone info, please contact the Malaria Research and Reference Reagent Resource Center, ATCC

http://www.malaria.mr4.org/mr4pages/index.html

Seq primer: ADP.

FEATURES

source

Location/Qualifiers

1..645

/organism="Plasmodium yoelii yoelii"

/mol_type="mRNA"

/strain="17XL"

/db_xref="taxon:73239"

/clone="PYCML60"

/dev_stage="asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/clone_lib="PYBS"

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcytalline cellulose columns. Total RNA was isolated using the guanidinium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved

with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to Hybrizap arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-GAL4) was excised from the Hybrizap vector and plasmid DNA isolated."

BASE COUNT 266 a 50 c 177 g 152 t

Query Match 2.4%; Score 56.6; DB 12; Length 645;
Best Local Similarity 47.4%; Pred. No. 0.12; Indels 0; Gaps 0;
Matches 170; Conservative 0; Mismatches 189;

107 GATGAGGTAGCAAGGCTGAGGAATCAATGCTGACCAATCGTCATCAAGTAACAGAC 166
120 GATGATGATGACGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 179
167 CAAGCTATGTCTACCTTCACATGCGGACCACTATCATTTTACCAAGTGAAGTCTTAT 226
180 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAA 239
227 GACGCTATCATGCTGAGATTAATCTGATGATGATGATGATGATGATGATGATGATGAT 286
240 GAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 299
287 GATATTGTTAATGAGCTCAAGGCTGATGATGATGATGATGATGATGATGATGATGATGAT 346
300 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 359
347 TACCTTAAGGATGCTCCGACCGGATTAACGTCCTGATCAAAAGAGAAATCAATGACAA 406
360 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 419
407 AAACAAGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 465
420 GAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 478

RESULT 15

BM162732 747 bp mRNA linear EST 04-DEC-2001
LOCUS BM162732 PyBS Plasmodium yoei11 cDNA clone PYCKW24 5' end,
DEFINITION mRNA sequence.

ACCESSION BM162732.1 GI:117308413

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@igir.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ARCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADP.

FEATURES

1. 747
Location/Qualifiers
/organism="Plasmodium yoei11 yoei11"
/mol_type="mRNA"
/strain="17XL"
/db_xref="taxon:73239"
/clone="PYCKW24"
/dev_stage="Asexual blood stages"

/lab host="E. coli XL-1 Blue"
/clone lib="PyBS"

/note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was collected from BAB/CD1 mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to Hybrizap arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-GAL4) was excised from the Hybrizap vector and plasmid DNA isolated."

BASE COUNT 330 a 58 c 195 g 164 t

Query Match 2.4%; Score 56.6; DB 12; Length 747;
Best Local Similarity 47.4%; Pred. No. 0.13; Indels 0; Gaps 0;
Matches 170; Conservative 0; Mismatches 189;

107 GATGAGGTAGCAAGGCTGAGGAATCAATGCTGACCAATCGTCATCAAGTAACAGAC 166
270 GATGATGATGACGACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 329
167 CAAGCTATGTCTACCTTCACATGCGGACCACTATCATTTTACCAAGTGAAGTCTTAT 226
330 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 389
227 GACGCTATCATGCTGAGATTAATCTGATGATGATGATGATGATGATGATGATGATGATGAT 286
390 GAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 449
287 GATATTGTTAATGAGCTCAAGGCTGATGATGATGATGATGATGATGATGATGATGATGAT 346
450 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 509
347 TACCTTAAGGATGCTCCGACCGGATTAACGTCCTGATCAAAAGAGAAATCAATGACAA 406
510 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 569
407 AAACAAGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 465
570 GAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 628

RESULT 16

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543

Email: bjoifus@icgr.org
 Clones are derived from the Entamoeba histolytica HM1:IMSS sheared DNA library
 Seq primer: M13-Reverse
 Class: shotgun
 High quality sequence start: 36
 High quality sequence stop: 816.
 Location/Qualifiers

FEATURES

1. 816
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HM1:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: pHOSt1 Site 1; Bat I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD. Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 125 a 265 c 54 g 372 t
 ORIGIN

Query Match 2.3%; Score 55.2; DB 28; Length 816;
 Best Local Similarity 42.0%; Pred. No. 0.27;
 Matches 324; Conservative 0; Mismatches 448; Indels 0; Gaps 0;

Oy 1425 ATGATGAATCCGCTAATAAGAAAATTGTTGATGATTTATGGCATTCTCAGACCAA 1484
 |||||
 Db 806 ACGATGATGAAGAAAGATGATGATGAAGATGATGATGATGATGATGATGATGATG 747
 1485 TTACCATCCAGAGGAGCTGGCAACCAATTCCTCAATTGAGTATCTGAGAGCAAG 1544
 |||||
 Db 746 AAGAGATGATGAAGATGATGATGAAGAGATGATGATGATGATGATGATGATGATG 687
 1545 TTCCGATTCCTCAATTAAGCTGATATACCACTCAGTGTATCAATTTTGTGTAAC 1604
 |||||
 Db 686 AAGATGATGATGAAGAGATGATGATGAAGAGATGATGATGAAGAGATGATGATG 627
 Oy 1605 ATGATATATACGATGATGAAGAGATGATGATTAACGCTCATATGCGCCATGTC 1664
 |||||
 Db 626 ATGACGATGATGATGAAGATGATGATGATGAAGAGATGATGATGATGATGATG 567
 1665 GGATTGGAAGATGATGCTTCTGATTAAGGAAAAGTTGACCTGACCTATATCTAAG 1724
 |||||
 Db 566 ACGATGAAGAGACGATGAAGAGACGATGAAGAGACGATGAAGAGACGATGAAG 507
 Oy 1725 AAAAGATATCTTACCTCCATCCAGCGCAGATGTTAAAGCAATCCAACTGGAG 1784
 |||||
 Db 506 ACGATGAAGAGACGATGAAGAGACGATGAAGAGACGATGAAGAGACGATGATG 447
 Oy 1785 GTGCAGACGCTATTTAATCTGTTGAAGAGGGGAAAAGCAATTCCTGCTGACTTC 1844
 |||||
 Db 446 AATTGGAATTGAAGACGATGATGAAGAGACGATGAAGAGACGATGATGAAG 387
 1845 CATATATGTTGAGATACAGTTGAGTTAAAGGTAATTGATTTATCTCTAAGG 1904
 |||||
 Db 386 ACGATGAAGAGACGATGAAGAGACGATGAAGAGACGATGAAGAGACGATGAAG 327
 Oy 1905 ATCATATCCATATATTAATTTGCTGTTGATGATGATGATGATGATGATGATG 1964
 |||||
 Db 326 ACGATGATGAAGAGACGATGATGAAGAGACGATGATGAAGAGACGATGATGAAG 267
 Oy 1965 GCTATATCCTTGAGAGATTTGTTGGACGATTAACTAGCTAGAGACCCCTGAGCAG 2024
 |||||
 Db 266 ACGATGATGAAGAGACGATGATGAAGAGACGATGATGAAGAGACGATGATGAAG 207

Oy 2025 GTCCACATTTAATGATGATGGGCAATGCGCACTGATGTTAGGCAAGAACCC 2084
 |||||
 Db 206 ACGATGATGAAGAGACGATGATGAAGAGATGATGATGATGAAGAGACGATGAAG 147
 Oy 2085 ACGATGAAGATCCAAATPAGAACTTCAAGCGGATGAAGACCGTGAAGAAACCTG 2144
 |||||
 Db 146 ATGATGAAGAGACGATGAAGAGACGATGAAGAGACGATGAAGAGACGATGAAG 87
 Oy 2145 CTGAGCCGAGATGCTTCACTAGAGACTGAAAAGTGAAGCCCACTCA 2196
 |||||
 Db 86 ACGATGAAGAGACGATGAAGAGACGATGAAGAGACGATGAAGAGACG 35

RESULT 17

LOCUS BM163120 647 bp mRNA linear EST 04-DEC-2001
 DEFINITION EST565643 PyBS Plasmodium yoelii yoelii cDNA clone PYCLB13 5' end,
 mRNA sequence.

ACCESSION BM163120 GI:17308801
 VERSION 1
 KEYWORDS EST.

SOURCE

ORGANISM Plasmodium yoelii yoelii
 Plasmodium yoelii yoelii
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 1 (bases 1 to 647)

REFERENCE

AUTHORS Carlton J.M., Daly T.M., Long C.A., Bergman L.W., Valdivia A.B.,
 Fraser C.M. and Carucci D.J.

TITLE

JOURNAL Unpublished
 COMMENT Contact: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlton@icgr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/index.html
 Seq primer: ADP.

FEATURES

source

Location/Qualifiers
 1. 647
 /organism="Plasmodium yoelii yoelii"
 /mol_type="mRNA"
 /strain="17XL"
 /db_xref="taxon:73239"
 /clone="PYCLB13"
 /dev_stage="asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /clone_lib="PyBS"
 /note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybriZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-GAL4) was excised from the HybriZAP vector and plasmid DNA isolated."

BASE COUNT 273 a 53 c 190 g 131 t
 ORIGIN

Query Match 2.3%; Score 55; DB 12; Length 647;
 Best Local Similarity 47.1%; Pred. No. 0.27;

Matches 169; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

QY 107 GATGAGTTAGCAGAGCGTGAAGAAATCAATGCTGAGCAATATGTCATCAAGATAACAGAC 166
 DB 238 GATGATGATGACGACGATGAAGATGACGATGAAGACGATGATGACGACGATGAC 297
 QY 167 CAAGCTATGTCACCTTACATGCGGACCACTATCATTATTACAAATGTAAGTTCCTTAT 226
 DB 238 GACGATGATGACGATGATGACGATGACGATGATGATGATGATGATGATGATGATG 357
 QY 227 GACGCTATGTCACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 286
 DB 338 GAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 417
 QY 287 GATATTGTTAATGAGGTCAAGGTGATGATGATGATGATGATGATGATGATGATGATG 346
 DB 418 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 477
 QY 347 TACCTTAAGATGCTGCCCGCGGATTAACGTCCTGTAACAAAGAGAAATCAATGACAA 406
 DB 478 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 537
 QY 407 AAACAAGACATGATCAACATGTCGAAGTGAATCCCAAGAAAGATGCTGCTTGC 465
 DB 538 GAGGATGAAGATGATGAAGAAAGATGACGATGATTTTAAAGAAAGATGATCAATGC 596

RESULT 18

BM160252

LOCUS BM160252 671 bp mRNA linear EST 04-DEC-2001
 DEFINITION EST562775 PyBS plasmodium yoelii yoelii cDNA clone PYCJ090 5' end,
 mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Plasmodium yoelii yoelii
 Plasmodium yoelii yoelii
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 1 (bases 1 to 671)
 Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Valdiva, A.B.,
 Fraser, C.M., and Carucci, D.J.
 Plasmodium yoelii EST project at TIGR
 Unpublished
 Contact: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlton@tigr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mt4.org/mr4pages/index.html
 Seq primer: ADF.

FEATURES

source

1..671
 /location/Qualifiers
 /organism="Plasmodium yoelii yoelii"
 /mol_type="mRNA"
 /strain="17XL"
 /db_xref="taxon:73239"
 /clone="PYCJ090"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /clone_lib="PyBS"
 /note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was
 collected from BAB6/cby mice infected with Py17XL
 parasites, and leukocytes removed by passage over
 microcrystalline cellulose columns. Total RNA was
 isolated using the guanidium isothiocyanate method, and
 mRNA isolated using oligo(dT)-cellulose chromatography.
 First strand cDNA synthesis was completed using a 50-base
 primer and reverse transcriptase in the presence of
 5-methyl dCTP. After second strand synthesis, uneven

termini were treated with Pfu DNA polymerase and EcoRI
 adaptors ligated to the blunt ends. The sample was cleaved
 with XhoI and separated on a Sephacryl S-500 column.
 Size-fractionated cDNA was precipitated and ligated to
 HybridZAP arms directionally using EcoRI-XhoI cleaved arms.
 After packaging, the phagemid vector (pAD-GAL4) was
 excised from the HybridZAP vector and plasmid DNA
 isolated."

BASE COUNT

288 a 50 c 184 g 149 t

Query Match 2.3%; Score 55; DB 12; Length 671;
 Best Local Similarity 47.1%; Pred. No. 0.27;

Matches 169; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

QY 107 GATGAGTTAGCAGAGCGTGAAGAAATCAATGCTGAGCAATATGTCATCAAGATAACAGAC 166
 DB 194 GATGATGATGACGACGATGAAGATGACGATGAAGACGATGATGACGACGATGAC 253
 QY 167 CAAGCTATGTCACCTTACATGCGGACCACTATCATTATTACAAATGTAAGTTCCTTAT 226
 DB 254 GACGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 313
 QY 227 GACGCTATGTCACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 286
 DB 314 GAGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 373
 QY 287 GATATTGTTAATGAGGTCAAGGTGATGATGATGATGATGATGATGATGATGATGATG 346
 DB 374 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 433
 QY 347 TACCTTAAGATGCTGCCCGCGGATTAACGTCCTGTAACAAAGAGAAATCAATGACAA 406
 DB 434 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 493
 QY 407 AAACAAGACATGATCAACATGTCGAAGTGAATCCCAAGAAAGATGCTGCTTGC 465
 DB 494 GAGGATGAAGATGATGAAGAAAGATGACGATGATTTTAAAGAAAGATGATCAATGC 552

RESULT 19

BM168242

LOCUS BM168242 694 bp mRNA linear EST 04-DEC-2001
 DEFINITION EST570765 PyBS plasmodium yoelii yoelii cDNA clone PYCJ090 5' end,
 mRNA sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Plasmodium yoelii yoelii
 Plasmodium yoelii yoelii
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 1 (bases 1 to 694)
 Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Valdiva, A.B.,
 Fraser, C.M., and Carucci, D.J.
 Plasmodium yoelii EST project at TIGR
 Unpublished
 Contact: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlton@tigr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mt4.org/mr4pages/index.html
 Seq primer: ADF.

FEATURES

source

1..694
 /location/Qualifiers
 /organism="Plasmodium yoelii yoelii"
 /mol_type="mRNA"
 /strain="17XL"
 /db_xref="taxon:73239"

```

/clone="PYCOX60"
/dev stage="Asexual blood stages"
/lab host="E. coli XL-1 Blue"
/clone.lib="PyBS"
/note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/cByJ mice infected with Py17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidium isothiocyanate method, and
mRNA isolated using oligo(dT)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, uneven
termini were treated with Pfu DNA polymerase and EcoRI
adaptors ligated to the blunt ends. The sample was cleaved
with XhoI and separated on a Sephacryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to
HybridZAP arms directionally using EcoRI-XhoI cleaved arms.
After packaging, the phagemid vector (pAD-GAL4) was
excised from the HybridZAP vector and plasmid DNA
isolated."

```

BASE COUNT 300 a 50 c 187 g 157 t

Query Match 2.3%; Score 55; DB 12; Length 694;
 Best Local Similarity 47.1%; Pred. No. 0.28;
 Matches 169; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

```

QY 107 GATGAGTTAGCAAGCGTGAAGAAATCAATGCTGCAAAATGTCATCAAGATAACAGAC 166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 192 GATGATGATGACGACGATGAAGATGACATGATGAAGATGATGATGATGATGATGATGAT 251
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 167 CAAGCTATGTCACCTTCACATGGCGACCATATCTTTTCAATGTAAGTTCCTAT 226
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 252 GACGATGATGACGATGATGACGATGACGATGATGATGATGATGATGATGATGATGATGAA 311
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 227 GACGCTATCATCAGTGAAGAAATTAATCTCATGAAAGATCCAACTATTAAGCTAAAGATGAG 286
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 312 GAAGATGACGATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 371
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 287 GATATTGTTAATGAGCTCAAGGCGATATGTTTATCAAGATGATGATGATGATGATGATGAT 346
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 372 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 431
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 347 TACCTTAAGATGCTGCCACGCGATTAACGTCCTACCAAAAGAGAAATCAATGACAA 406
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 432 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 407 AAACAAGACATAGTCAACATCGTGAAGTGAATCCCAAGAAACGATGCTGCTTGC 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 492 GAGGATGAAGATGATGAAGAAAGATGACGATGATTTCTAAAAAAGAGGTGATCAATGC 550
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 20
 BM160500 717 bp mRNA linear EST 04-DEC-2001
 LOCUS EST563023 PyBS plasmodium yoelii yoelii cDNA clone pYCS13 5' end,
 DEFINITION mRNA sequence.

ACCESSION BM160500.1 GI:17306181
 VERSION BM160500
 KEYWORDS EST.
 SOURCE Plasmodium yoelii yoelii
 ORGANISM Plasmodium yoelii yoelii
 Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.

REFERENCE 1 (bases 1 to 717)
 AUTHORS Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Vaidya, A.B.,
 Fraser, C.M., and Carucci, D.J.

TITLE Plasmodium yoelii EST project at TIGR
 JOURNAL Unpublished
 COMMENT Contact: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlone@igr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/index.html
 Seq primer: ADF.

FEATURES

source

```

1. 717
   location/Qualifiers
   /organism="Plasmodium yoelii yoelii"
   /mol_type="mRNA"
   /strain="17XL"
   /db_xref="taxon:73239"
   /clone="PYCS13"
   /dev stage="Asexual blood stages"
   /lab host="E. coli XL-1 Blue"
   /clone.lib="PyBS"

```

/note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybridZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (pAD-GAL4) was excised from the HybridZAP vector and plasmid DNA isolated."

BASE COUNT 311 a 52 c 193 g 161 t

Query Match 2.3%; Score 55; DB 12; Length 717;
 Best Local Similarity 47.1%; Pred. No. 0.28;
 Matches 169; Conservative 0; Mismatches 190; Indels 0; Gaps 0;

```

QY 107 GATGAGTTAGCAAGCGTGAAGAAATCAATGCTGCAAAATGTCATCAAGATAACAGAC 166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 205 GATGATGATGACGACGATGAAGATGACATGATGAAGATGATGATGATGATGATGATGATGAT 264
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 167 CAAGCTATGTCACCTTCACATGGCGACCATATCTTTTCAATGTAAGTTCCTAT 226
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 265 GACGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAA 324
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 227 GACGCTATCATCAGTGAAGAAATTAATCTCATGAAAGATCCAACTATTAAGCTAAAGATGAG 286
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 325 GAAGATGACGATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 384
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 287 GATATTGTTAATGAGCTCAAGGCGATATGTTTATCAAGATGATGATGATGATGATGATGATGAT 346
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 385 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 444
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 347 TACCTTAAGATGCTGCCACGCGATTAACGTCCTACCAAAAGAGAAATCAATGACAA 406
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 445 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 504
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 407 AAACAAGACATAGTCAACATCGTGAAGTGAATCCCAAGAAACGATGCTGCTTGC 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 505 GAGGATGAAGATGATGAAGAAAGATGACGATGATTTCTAAAAAAGAGGTGATCAATGC 563
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 21
 BM162492 756 bp mRNA linear EST 04-DEC-2001
 LOCUS EST565015 PyBS Plasmodium yoelii yoelii cDNA clone pYCR28 5' end,
 DEFINITION mRNA sequence.
 ACCESSION BM162492
 VERSION BM162492.1 GI:17308173

77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 228 a 206 c 86 g 303 t

Query Match 2.3%; Score 55; DB 28; Length 823;
Best Local Similarity 46.7%; Pred. No. 0.29;
Matches 175; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

223 TTATGACGCTATCATCAGTGAAGATTACTCATGAAGATCCAACTATAAGCTAAAGA 282
223 TTATGATGAATGCAATTAGAGAAATGTTACGAAAGAAATACCTGGAATACGA 664
283 TGAAGATATTGTTAATGAGGTCAAGGTGATATGTTATCAAGGTAGATGAAATACCTA 342
663 AGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 604
343 TGTTCCTTAAGAGATGCTGCCACCGGATTAACCTCCGTAACAAAGAGAAATCAATG 402
603 CGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 544
403 ACAAACAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 462
543 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 484
463 TGCCTTGGACGCTTCCAAAGACGCTATACATGATGATGATGATGATGATGATGATG 522
483 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 424
523 TGTATCATAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 582
423 CGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 364
583 CATTCCTTAAGATGA 597
363 CGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 349

RESULT 25 843 bp DNA linear GSS 07-AUG-2001
BH139532
LOCUS
DEFINITION ENTNG88TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
ACCESSION BH139532
VERSION BH139532.1 GI:15098593
KEYWORDS GSS.

SOURCE Entamoeba histolytica
ORGANISM Entamoeba histolytica
REFERENCE 1 (bases 1 to 843)
AUTHORS Loftus, B., Wang, Z., Van Aken, S. and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library (2001)
COMMENT Unpublished
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: b.loftus@tigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
Seq primer: M13-Forward
Classes: Shotgun
High quality sequence start: 4
High quality sequence stop: 714.
Location/Qualifiers

source

1. 843
/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHO1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaubin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 324 a 82 c 205 g 232 t

Query Match 2.3%; Score 55; DB 28; Length 843;
Best Local Similarity 46.7%; Pred. No. 0.3;
Matches 175; Conservative 0; Mismatches 200; Indels 0; Gaps 0;

223 TTATGACGCTATCATCAGTGAAGATTACTCATGAAGATCCAACTATAAGCTAAAGA 282
159 TTATGATGAATGCAATTAGAGAAATGTTACGAAAGAAATACCTGGAATACGA 218
283 TGAAGATATTGTTAATGAGGTCAAGGTGATATGTTATCAAGGTAGATGAAATACCTA 342
219 AGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 278
343 TGTTCCTTAAGAGATGCTGCCACCGGATTAACCTCCGTAACAAAGAGAAATCAATG 402
279 CGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 338
403 ACAAACAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 462
339 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 398
463 TGCCTTGGACGCTTCCAAAGACGCTATACATGATGATGATGATGATGATGATGATG 522
399 TGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 458
523 TGTATCATAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 582
459 CGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 518
583 CATTCCTTAAGATGA 597
519 CGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 533

RESULT 26 721 bp mRNA linear EST 03-NOV-2002
BH141179/c
LOCUS
DEFINITION BW141179 Nori Satoh unpublished cDNA library, gastrula and neurula
ACCESSION BW141179
VERSION BW141179.1 GI:24498404
KEYWORDS EST.
SOURCE Ciona intestinalis
ORGANISM Ciona intestinalis
REFERENCE 1 (bases 1 to 721)
AUTHORS Satou, Y., Shin-i, T., Kohara, Y. and Satoh, N.
TITLE Expressed genes in Ciona intestinalis (2002c)
JOURNAL Unpublished
COMMENT Contact: Nori Satoh
Department of Zoology

Phlebobranchia; Clonidae; Ciona.
Chordata; Urochordata; Ascidiacea; Enterogona;

Kyoto University
Sakyo-ku, Kyoto, 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoch@ascidian.zool.kyoto-u.ac.jp

FEATURES
Source

```
/organism="Cliona intestinalis"  
/mol_type="mRNA"  
/db_xref="taxon:7719"  
/clone="rcign058f11"  
/tissue_type="whole body"  
/dev_stage="gastrea and neurula"  
/clone_lib="Nori Satoh unpublished cDNA library, gastrula and neurula"
```

	a	c	g	t	others
BASE COUNT ORIGIN	150	209	73	287	2

Query Match	2.3%	Score 53.8;	DB 13;	Length 721;
Best Local Similarity	44.1%;	Pred. No. 0.51;		
Matches 223; Conservative	0;	Mismatches 283;	Indels 0;	Gaps 0;

Qy	12	TCGAAGATCAATGCTGAGCAAAATCGTCTCAAGATAACAGCCAAAGCTTGTGACTTC	183
Db	688	TGACGAAGATGATGNGCAGAGATGACGAAATATATACGAAGATGATGACGAAATGATGA	629
Qy	184	ACATGCGCACCACTATCATTTATTTCAATGTAAGTTCTCTATGACGCTATCATCACTGA	243
Db	628	CGAAGATGATGACGAAGATGACGAAATATGATCAGCAAGATGATGACGAAAGTATGACGA	569
Qy	244	AGATTTACTCATGAAGAATCCAAACCTATTAAGCTAAAGATGAGATATTTGTTATGAGT	303
Db	568	AGATGATGACGAAGATGACGAAATATGATGACGAAGATATGATCGAAAGTATGACGAGA	509
Qy	304	CAAAGGTGATATGTTTATCAGGTGATGAGAAATATCTATGTTTACTTTAAGGTGCTGC	363
Db	508	TGATGACGAAGATGATGAAGAAAGATGATGACGTAGATATGATATACGAAGATGATGACGT	449
Qy	364	CCACGCGGATTAACGTCCGTACAAAGAGGAAATCAATGACGAAAGAACAGAGCTACTGA	423
Db	448	AGATGATGATGACGAAGATGATGACGAAAGATGATATACGAAATATGATGACGATGATGA	389
Qy	424	ACATCGTAGAGTGAACCTCCAAGAAACGATGTCGTGTCCTTGCGACGGTTCGCAAG	483
Db	388	CGATGATGACGAAGATGACGAAGATGATGAGAAAGATGATGACGTAAGATGATGACGTAGA	329
Qy	484	AGCCTTACTACAGATGATGCTTATATCTTTAATGCTTCTGATATCATAGAGGTACTGG	543
Db	328	TGATGATGACGAAGATGATGACGAAGATGATGACCTAGATATATATGACGAAATGATGA	269
Qy	544	TGATGCTTATATCGTTCCCTCATGAGATCATTTACCATTAACCTCTAAGATGAGTATC	603
Db	268	CGAAGATGATGACGAAGATGATGACGAAGATGATGACGAAGATGATGATGACGATGACGTAGA	209
Qy	604	AGCTACCGAGTTGCTGCTGCAAG	629
Db	208	TGATGCTGACGAAGATGATGACAAAG	183

RESULT	27
LOCUS	BM144040/c
DEFINITION	BM144040 656 bp mRNA linear EST 03-NOV-2002
ACCESSION	BM144040
VERSION	BM144040
KEYWORDS	Clona intestinalis cDNA clone rc1gn057b16 3', mRNA sequence.
SOURCE	BM144040.1 GI:24501265
ORGANISM	EST.
REFERENCE	Clona intestinalis
AUTHORS	Clona intestinalis Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona; Phlebobranchia; Cloniidae; Clona. 1 (bases 1 to 656) Satou,Y., Shin-I,T., Kohara,Y. and Satoh,N.

TITLE	Expressed genes in <i>Ciona intestinalis</i> (2002c)
JOURNAL	unpublished
COMMENT	Contact: Nori Satoh

FEATURES
SOURCE

Email: satoh@ascidian.zool.kyoto-u.ac.jp.

```

      /clone_lib="Nor1 Sacch unpublished cDNA library, gastrula
      and neurula"
BASE COUNT      145 a      185 c      66 g      260 t
ORIGIN

```

Query Match	2.2%	Score 53.6;	DB 13;	Length 656;
Best Local Similarity	45.0%;	Pred. No. 0.55;		
Matches 200; Conservative	0;	Mismatches 244;	Indels 0;	Gaps 0;

QY	186	ATGGGACACCATATCATTTATTATACATGGTAAGGTTCCCTTATGACGCTATCATCAGTGAAG	245
Db	651	AGATGATACCGAAGTACGAAGTAATGTACGAAAGTGAATGACGAAAGTGAAGTACGAAAG	592
QY	246	AATTACTGTAAGAAATCCAACTATTAAAGCTAAAGAATGATAGGATATTGTTAATAGGTCA	305
Db	591	ATGATGACCAAGATACCAAAAATGATGTCGAAGATGTATGACGAAGTGATGACGAAGATG	532
QY	306	AGGGTGATATGTTATCAAGGTAGTGAAGAAATCTATGTTTACCTTAAGGATGCTGCC	365
Db	531	ATGACCAAGATATGAAAGAGTGAATGACGTAGTAATGATGACGAAGATGATGACCTAG	472
QY	366	ACGCGGATACGTCCTCGTCAAAAAGAGAAATCAATGCACAAAAACAAGACATAGTCAC	425
Db	471	ATGATGATGACGAAGATGATGACGAAGTATGATGATGACGAAAATGATGACGATGATGACG	412
QY	426	ATCTGTAAGGTGAACCTCCAAAACCAATCGATGCTGCTGTTGCTTGACAGCTTCCGAAGAC	485
Db	411	ATGATGACCAAGATGACGAAGATGTATGTAAGAAAGTATGATGACGTATGATGATGACGTATGATG	352
QY	486	GCATATCTACAGATGATGGTTATATCTTTAATGCTTCGTGATTCATGTAGAGGATAGTG	545
Db	351	ATGATGACCAAGATATATGACGAAGTATGATGACGTATGATGATGACGAAATGATGATGACG	292
QY	546	ATGCTTATATCGTTCCTCATGAGATCTTACCATTTACATTTCTTAAGATGAGTATACG	605
Db	291	AGATGATGACGAAGATGATGACGAAGATGACGAAGATGATGACGAAGATGACGTAGATG	232
QY	606	CTAGCGAGTTGGCTGCTGCAGAG	629
Db	231	ATGCTGACGAAGATGATGACAA	208

RESULT 28							
LOCUS	844 bp	DNA	linear	GSS	13-MAR-2002		
DEFINITION	BX139987 genomic clone DKEX-99E7, genomic survey sequence.						
ACCESSION	BX139987						
VERSION	BX139987.1 GI:27971314						
KEYWORDS	GSS.						
SOURCE	Danio rerio (zebrafish)						
ORGANISM	Danio rerio						
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes; Cyprinidae; Danio.						
REFERENCE	1 (bases 1 to 844)						
AUTHORS	Humphray,S.J., Huckle,E. and Durham,J.L.						

TITLE Direct Submission
JOURNAL Submitted (13-MAR-2003) The Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humnuey@sanger.ac.uk Unpublished
COMMENT This sequence was generated from the T7 end of BAC 9957. 9957 is part of the Daniokey BAC Library created by R. Plasterk and N.V. Keygene. Further details: http://www.sanger.ac.uk/Projects/D_reio/.
FEATURES
source Location/Qualifiers
1..844
/organism="Danio rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKEY-9957"
/tissue="Testis"
/note="Vector pIndigoBAC-536"
BASE COUNT 406 a 49 c 81 g 308 t
ORIGIN
Query Match 2.2%; Score 53.6; DB 29; Length 844;
Best Local Similarity 45.6%; Pred. No. 0.6;
Matches 188; Conservative 0; Mismatches 224; Indels 0; Gaps 0;
OY 186 ATGGGACCATATCATTTATTTCAATGGTAGGTCCTTATGACGCTATCATGAG 245
DB 417 ATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 476
OY 246 AATTACTGATGAAGATCCAACTTAAGCTAAAGTAGAGATATTGTTATGAGCTCA 305
DB 477 ATTAATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 536
OY 306 AGGGTGATATGTTATCAAGTAGATGGAATACTATGTTTACCTTAAGATGCTGCC 365
DB 537 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 596
OY 366 ACGCGGATPACCTCCGTACAAAAGAGAAATTAATGACAAAACAAGCATAGTCAAC 425
DB 597 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 656
OY 426 ATTCGTAAGGTGGAAGCTCCAGAAACGATGCTGCTGCTCCCTGGCAGCTTGCAAGAC 485
DB 657 ATTAATGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 716
OY 486 GCTATTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 545
DB 717 ACAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 776
OY 546 ATGCTTAATATCGTTCCTCATGAGATCATTAACATTAATTAATTAATTAATTA 597
DB 777 ATACTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 828
RESULT 29
CD099026 752 bp mRNA linear EST 15-MAY-2003
LOCUS AGENCOURT_14011375 NICHD XGC Tad1 Xenopus laevis cDNA clone
DEFINITION IMAGE:6939285 5', mRNA sequence.
ACCESSION CD099026 GI:30752129
VERSION CD099026.1 GI:30752129
KEYWORDS EST.
SOURCE Xenopus laevis (African clawed frog)
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus.
REFERENCE 1 (bases 1 to 752)
AUTHORS NCI-CCAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Drs. Donald Brown and Liqun Cai

CDNA Library Preparation: CLONTech
JOURNAL cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: NCI-CCAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LICM3263 row: n column: 20
High quality sequence stop: 496.
FEATURES
source Location/Qualifiers
1..752
/organism="Xenopus laevis"
/mol_type="mRNA"
/db_xref="taxon:8355"
/clone="IMAGE:6939285"
/dev_stage="metamorphosis stage 53"
/clone_lib="NICHD XGC Tad1"
/note="Organ: Developing Tadpole; Vector: pDNR-LIB; Site 1: Sfi; Site 2: Sfi; 5' and 3' adaptors were used in cloning as follows: 5' adaptor sequence: 5'-CACGGCCATTTAGGCC-3' and 3' adaptor sequence: 5'-ATCTAGAGCGCGCCGCGACATG-dt(30)BN-3' (where B = A, C, or G and N = A, C, G, or T). Average insert size 1.6 kb (range 0.9-3.0 kb). 15/15 colonies contained inserts by PCR. This library was enriched for full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA)."
BASE COUNT 247 a 128 c 208 g 169 t
ORIGIN
Query Match 2.2%; Score 53.2; DB 14; Length 752;
Best Local Similarity 47.8%; Pred. No. 0.7;
Matches 154; Conservative 0; Mismatches 168; Indels 0; Gaps 0;
OY 107 GATGAGTTAGCAAGCGGAGGAATCAATGCTGACAAATGCTATCAAGTTACAGC 166
DB 348 GACGATGATGACGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 407
OY 167 CAAGCTATGTCACCTTCACATGCGACCACTATCTTATTAATGATGATGATGATG 226
DB 408 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 467
OY 227 GACGCTATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 286
DB 468 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 527
OY 287 GATATTTGTTAATGAGCTCAAGGTGATGATGATGATGATGATGATGATGATGAT 346
DB 528 GACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 587
OY 347 TACCTTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 406
DB 588 AACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 647
OY 407 AAACAGAGCATAGTCAATC 428
DB 648 CACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 669
RESULT 30
BW273704 671 bp mRNA linear EST 11-NOV-2002
LOCUS BW273704 Nori Satoh unpublished cDNA library, gastrula and neurula
DEFINITION BW273704 Nori Satoh unpublished cDNA library, gastrula and neurula
ACCESSION BW273704 GI:24854315
VERSION BW273704.1 GI:24854315
KEYWORDS EST.
SOURCE Ciona intestinalis
ORGANISM Ciona intestinalis
Eukaryota; Metazoa; Chordata; Urochordata; Ascidiacea; Enterogona;
Phlebobranchia; Cionidae; Ciona.
REFERENCE 1 (bases 1 to 671)
AUTHORS Satou, Y., Shin-I, T., Kohara, Y. and Satoh, N.
TITLE Expressed genes in Ciona intestinalis (2002c)
JOURNAL Unpublished

COMMENT Contact: Nori Satoh
Department of Zoology
Kyoto University
Sakyo-ku, Kyoto 606-8502, Japan
Tel: 81-75-753-4081
Fax: 81-75-705-1113
Email: satoh@asc.idian.zool.kyoto-u.ac.jp.
Location/Qualifiers

FEATURES
source
1..671
/organism="Clona intestinalis"
/mol_type="mRNA"
/db_xref="taxon:7719"
/clone_id="c1gn067b16"
/tissue_type="whole body"
/dev_stage="gastrula and neurola"
/clone_lib="Nori Satoh unpublished cdna library, gastrula and neurola"

BASE COUNT 272 a 67 c 182 g 147 t 3 others
ORIGIN

Query Match 2.2%; Score 52.8; DB 13; Length 671;
Best Local Similarity 44.8%; Pred. No. 0.83; Mismatches 240; Indels 0; Gaps 0;
Matches 195; Conservative 0;

QY 195 ACTATCATTTATACAAAGTGAAGTTCCTTATGACGCTATCATCGAGAGATTACTCA 254
DB 2 ACGAAGATGACGAAATGATGACGAAGATGATGACGAAGATGATGACGAAGATGATGACG 61
QY 255 TGAAGATCCAACTATTAAGCTAAAAGATGAGATATTGTTAATAGGTCAGAGGTGAT 314
DB 62 AAGATGACGAAATGATGACGAAGATGATGACGAAGATGATGACGAAGATGATGACGAAG 121
QY 315 ATGTTATCAAGGTAGTGAAGAAATCTATGTTTACCTTAAGGATCTGCCCGCGGATA 374
DB 122 ATGATGAAGAGATGATGACGATGATGATGATGATGATGATGATGATGATGATGATG 181
QY 375 ACGTCCGTAACAAGAGGAAATCAATCGACAAACAAAGAGCATGTCACATCGTGAAG 434
DB 182 ACGAAGATGATGACGAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 241
QY 435 GTGAACTCCAAAGAACGATGCTGCTGTTGCTTGGACGTTGCCAAGAGCGCTTACTA 494
DB 242 AAGATGACGAAGATGATGAAAGATGATGACGAGATGATGATGATGATGATGATGACG 301
QY 495 CAGATGATGTTATATCTTTATGCTTCTGATATATAGAGATACCTGATGCTTATA 554
DB 302 AAGATGATGACGAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 361
QY 555 TCGTTCCTCATGAGATCATTTACATTCCTTAAGATGATGATGATGATGATGATGATG 614
DB 362 ACGAAGATGATGACGAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 421
QY 615 TGGCTCTGCAGAG 629
DB 422 AAGATGATGACAAAG 436

RESULT 31
A2530768 890 bp DNA linear GSS 03-NOV-2000
LOCUS ENTBHS5TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
DEFINITION genomic, genomic survey sequence.
ACCESSION A2530768
VERSION A2530768.1 GI:11084894
KEYWORDS GSS.
SOURCE Entamoeba histolytica
ORGANISM Entamoeba histolytica
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 890)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
Unpublished

COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjl@fuserc1gr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 20
High quality sequence stop: 822.
Location/Qualifiers

FEATURES
source
1..890
/organism="Entamoeba histolytica"
/mol_type="genomic DNA"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pHO51; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

BASE COUNT 375 a 61 c 225 g 229 t
ORIGIN

Query Match 2.2%; Score 52.8; DB 28; Length 890;
Best Local Similarity 45.1%; Pred. No. 0.91; Mismatches 237; Indels 0; Gaps 0;
Matches 195; Conservative 0;

QY 198 ATCATTTATACATGTAAGTTCCTTATGACGCTATCATCGAAGATTACTCATGA 257
DB 298 AACTTGAAGAGATTATCACTAGATGATGATGATGATGATGATGATGATGATGATGATG 357
QY 258 AAGATCCAACTATTAAGCTAAAAGATGAGATATTGTTAATGAGTCAAGGCTGATATG 317
DB 358 ATATGTAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 417
QY 318 TTATCAAGTATGATGAAATATCTATGTTTACCTTAAGATGCTGCCACCGGATACG 377
DB 418 AAGATGATGATTTGAAGAAAGACATGATGATGATGATGATGATGATGATGATGATGATG 477
QY 378 TCCGTACAAAAGAGAAATCAATGACAAAAAAGAGCATGATCAACATGTGAAGTG 437
DB 478 ATGAAGAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 537
QY 438 GAATCCAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 497
DB 538 ATGATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 597
QY 498 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 557
DB 598 ACGATGATGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 657
QY 558 TTCCCTCATGAGATCATTTACATTCCTTAAGATGATGATGATGATGATGATGATGATG 617
DB 658 ACGATGATGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 717
QY 618 CTGCTGCAGAG 629
DB 718 ATGAAGAGATG 729

RESULT 32
 LOCUS BM161314 677 bp mRNA linear EST 04-DEC-2001
 DEFINITION mRNA sequence.
 ACCESSION BM161314
 VERSION BM161314.1 GI:17306995
 KEYWORDS EST.
 SOURCE Plasmodium yoelii yoelii
 ORGANISM Plasmodium yoelii yoelii
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 677)
 Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
 Fraser,C.M. and Carucci,D.J.
 Plasmodium yoelii EST project at TIGR
 TITLE Unpublished
 JOURNAL
 COMMENT Contact: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlton@tigr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/index.html
 Seq primer: ADP.

FEATURES
 source
 1..677
 /organism="Plasmodium yoelii yoelii"
 /mol_type="mRNA"
 /strain="17XL"
 /db_xref="taxon:73239"
 /clone="PYCND34"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /clone_1lb="PyBS"
 /note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was
 collected from BALB/cByJ mice infected with Py17XL
 parasites, and leukocytes removed by passage over
 microcrystalline cellulose columns. Total RNA was
 isolated using the guanidium isothiocyanate method, and
 mRNA isolated using oligo(dT)-cellulose chromatography.
 First strand cDNA synthesis was completed using a 50-base
 primer and reverse transcriptase in the presence of
 5-methyl dCTP. After second strand synthesis, uneven
 termini were treated with Pfu DNA polymerase and EcoRI
 adaptors ligated to the blunt ends. The sample was cleaved
 with XhoI and separated on a Sephacryl S-500 column.
 Size-fractionated cDNA was precipitated and ligated to
 HybridZAP arms directionally using EcoRI-XhoI cleaved arms.
 After packaging, the phagemid vector (PAD-GAL4) was
 excised from the HybridZAP vector and plasmid DNA
 isolated."

BASE COUNT 288 a 50 c 186 g 153 t
 ORIGIN

Query Match 2.2%; Score 52.4; DB 12; Length 677;
 Best Local Similarity 48.3%; Pred. No. 1;
 Matches 146; Conservative 0; Mismatches 156; Indels 0; Gaps 0;
 164 GACCAAGGTATGTCATTACATGGCGACCACTATCATATATACAGTGAAGTTCT 223
 DB 250 GACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 309
 QY 224 TATGACCTATCATCATGAGATTAATCTCATGAAGATCAAACTATAAGTTAAAGAT 283
 DB 310 GAAGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 369
 QY 284 GAGGATATTGTTATAGGTCAGAGGTGATATGTTATCAAGGTAGTGAATAATCTAT 343
 DB 370 GATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 429

QY 344 GTTTACCTTAAGAGTCTGCCACCGGATTAACGTCCGTACAAAGAGAAATCAATCGA 403
 DB 430 GATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 489
 QY 404 CAAAAACAAGACATATGTCATCATGCTGAGGTGAACTCAAGAAACGATGCTGT 463
 DB 490 GATGAGATGAAGATGATGATGAAGATGATGATGATGATGATGATGATGAT 549
 QY 464 GC 465
 DB 550 GC 551

RESULT 33
 LOCUS BM169486 725 bp mRNA linear EST 04-DEC-2001
 DEFINITION ESTS72009 PyBS Plasmodium yoelii yoelii cDNA clone PYCPR57 5' end,
 mRNA sequence.
 ACCESSION BM169486
 VERSION BM169486.1 GI:17302718
 KEYWORDS EST.
 SOURCE Plasmodium yoelii yoelii
 ORGANISM Plasmodium yoelii yoelii
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 725)
 Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
 Fraser,C.M. and Carucci,D.J.
 Plasmodium yoelii EST project at TIGR
 TITLE Unpublished
 JOURNAL
 COMMENT Contact: Jane Carlton
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-530-9319
 Fax: 301-838-0208
 Email: carlton@tigr.org
 For clone info, please contact the Malaria Research and Reference
 Reagent Resource Center, ATCC
 http://www.malaria.mr4.org/mr4pages/index.html
 Seq primer: ADP.

FEATURES
 source
 1..725
 /organism="Plasmodium yoelii yoelii"
 /mol_type="mRNA"
 /strain="17XL"
 /db_xref="taxon:73239"
 /clone="PYCPR57"
 /dev_stage="Asexual blood stages"
 /lab_host="E. coli XL-1 Blue"
 /clone_1lb="PyBS"
 /note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was
 collected from BALB/cByJ mice infected with Py17XL
 parasites, and leukocytes removed by passage over
 microcrystalline cellulose columns. Total RNA was
 isolated using the guanidium isothiocyanate method, and
 mRNA isolated using oligo(dT)-cellulose chromatography.
 First strand cDNA synthesis was completed using a 50-base
 primer and reverse transcriptase in the presence of
 5-methyl dCTP. After second strand synthesis, uneven
 termini were treated with Pfu DNA polymerase and EcoRI
 adaptors ligated to the blunt ends. The sample was cleaved
 with XhoI and separated on a Sephacryl S-500 column.
 Size-fractionated cDNA was precipitated and ligated to
 HybridZAP arms directionally using EcoRI-XhoI cleaved arms.
 After packaging, the phagemid vector (PAD-GAL4) was
 excised from the HybridZAP vector and plasmid DNA
 isolated."

BASE COUNT 263 a 64 c 156 g 242 t
 ORIGIN

Query Match 2.2%; Score 52.4; DB 12; Length 725;
 Best Local Similarity 48.3%; Pred. No. 1;
 Matches 146; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

Oy 164 GACCAAGGCTATGTCACATTCACATGCGACCACTATCATTTATTAACATGTAAGTTCCT 223
 Db 41 GACGACGATGATGACGATGATGACGATGACGATGATGATGATGATGATGATGATGATGAT 100
 Oy 224 TATGACGCTATCATCATGTAAGATTTACTCATGAAATTCCAACTATTAAGCTAAAGAT 283
 Db 101 GAAGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 160
 Oy 284 GAGGATATTTTAAAGAGTCAAGGTGATGATGATGATGATGATGATGATGATGATGATGAT 343
 Db 161 GATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 220
 Oy 344 GTTACCTTAAGATGCTGCCACCGGATTAACCTCCGTAACAAAGAGAAATCAATCGA 403
 Db 221 GATGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 280
 Oy 404 CAAAAACAAGACATAGTCAACATGCTGAAGGTGAACTCCAAAGAACGATGCTGCTT 463
 Db 281 GATGAGGATGAAGATGATGAAGAGATGACGATGATGATGATGATGATGATGATGATGAT 340
 Oy 464 GC 465
 Db 341 GC 342

RESULT 34

LOCUS A2527885 866 bp DNA linear GSS 03-NOV-2000
 DEFINITION ENTCA197R Entamoeba histolytica Sheared DNA Entamoeba histolytica
 genomic, genomic survey sequence.

ACCESSION A2527885
 VERSION A2527885.1 GI:11080056
 KEYWORDS GSS.

SOURCE Entamoeba histolytica

ORGANISM Entamoeba histolytica
 Eukaryota; Entamoebidae; Entamoeba.

REFERENCE 1 (bases 1 to 866)
 AUTHORS Loftus, B., Van Aken, S., and Fraser, C.
 TITLE Determination of clone end sequences from Entamoeba histolytica
 HM1:IMSS sheared DNA library
 JOURNAL Unpublished

COMMENT Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@tigr.org

Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
 DNA library
 Seq primer: M13-Reverse
 Class: Shotgun
 High quality sequence start: 15
 High quality sequence stop: 789.

FEATURES

source

1..866
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HM1:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: pHOSt1; Site 1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaudin and B.
 Batelli, Oxford University Press, 1999)."
 BASE COUNT 353 a 116 c 181 g 216 t
 ORIGIN

Query Match 2.2%; Score 52.4; DB 28; Length 866;
 Best Local Similarity 53.4%; Pred. No. 1.1;
 Matches 110; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

Oy 2004 ACCGTAGAACCCCGAGAACGCTCCACATCTATATGATGATGGGCAATGCCAGTGAGC 2063
 Db 17 ATGTAGGAAGAGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 76
 Oy 2064 ATGTGTTAGGCAAGAAAGACACAGTGAAGTCCAAATTAAGACTCAAGCGGATGAG 2123
 Db 77 ATGAAGACGACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 136
 Oy 2124 AGCCAGTAGAGAAACACCTGCTAGCCAGAGTCCCTCAAGTGAAGACTGAAAAAGTAG 2183
 Db 137 ATGAAGAAGACGATGAAGAACGATGATGATGATGATGATGATGATGATGATGATGAT 196
 Oy 2184 AAGCCCACTCAAGAACGAGAGTT 2209
 Db 197 ATGAAGAAGACGATGAAGAACGAT 222

RESULT 35

LOCUS BH149983 976 bp DNA linear GSS 27-AUG-2001
 DEFINITION ENTQD37R Entamoeba histolytica Sheared DNA Entamoeba histolytica
 genomic, genomic survey sequence.

ACCESSION BH149983
 VERSION BH149983.1 GI:15311165
 KEYWORDS GSS.

SOURCE Entamoeba histolytica

ORGANISM Entamoeba histolytica
 Eukaryota; Entamoebidae; Entamoeba.

REFERENCE 1 (bases 1 to 976)
 AUTHORS Loftus, B., Wang, Z., Van Aken, S., and Fraser, C.
 TITLE Determination of clone end sequences from Entamoeba histolytica
 HM1:IMSS sheared DNA library (2001)
 JOURNAL Unpublished

COMMENT Contact: Brendan J Loftus
 Department of Eukaryotic Genomics
 The Institute for Genomic Research
 9712 Medical Center Dr., Rockville, MD 20850, USA
 Tel: 301 838 0208
 Fax: 301 838 3543
 Email: b.loftus@tigr.org

Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
 DNA library
 Seq primer: M13-Forward
 Class: Shotgun
 High quality sequence start: 38
 High quality sequence stop: 740.

FEATURES

source

1..976
 /organism="Entamoeba histolytica"
 /mol_type="genomic DNA"
 /strain="HM1:IMSS"
 /db_xref="taxon:5759"
 /clone_lib="Entamoeba histolytica Sheared DNA"
 /note="Vector: pHOSt1; Site 1: Bst I; Constructed at The
 Institute for Genomic Research (TIGR), Rockville, MD.
 Genomic DNA isolated from broth cultures of E. histolytica
 using a method described by Clark and Diamond (Clark,
 C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
 method for isolate identification. Exp. Parasitol.
 77:450.). The DNA was mechanically sheared to give a
 tight size distribution (~2 kb). The v + i method used for
 the library construction is described in detail in Smith,
 H.O. and Venter, J.C. (Making small insert libraries for
 whole genome shotgun sequencing projects. In Genome

Sequencing: A Practical Approach, eds. M. Vaubin and B. Baretll, Oxford University Press, 1999. "

Query Match 2.2%; Score 52.4; DB 28; Length 976;
Best Local Similarity 45.7%; Pred. No. 1.1;
Matches 182; Conservative 0; Mismatches 216; Indels 0; Gaps 0;

106 TGATGAGGTAGCAAGCGTGAAGGATCAATGCTGCAAAATCGTCATCAATAGCA 165
166 CCAGGCTATGTCACCTTCACATGCGCCACCTATCATTTATTAAGTAAAGTTCTTA 225
384 TGATGAGAAGAGCGATGATGAAGAGACGATGATGAAGAGAGAGAGAGAGAGAG 443
226 TGAGCCTATCATCACTGAAGAAATTAATCTGAAAGATCCAACTTAAGCTTAAAGATGA 285
444 TGATGAGAAGAGCGATGATGAAGAGACGATGATGAAGAGAGAGAGAGAGATGA 503
286 GGATATTTGTTATGAGTCAAGGTCAGATGATTTATCAAGTATGAGAAATCTATGCT 345
504 TGATGAGAAGAGCGATGATGAAGAGATGATGAATTAATTAAGAGAGATGAAGATGA 563
346 TTACCTTAAGATGCTGCCACGCGATTAACCTCCGTAACAAAGAGAGAAATCAATGACA 405
564 TGATGAGAAGAGCGATGATGAAGAGATGATGAAGAGAGATGATGAAGAGAGAGAG 623
406 AAAACAAGAGCGATGATCAACATGCTGAAGGTGAAGCTCCAGAGAGAGATGCTGTTGC 465
624 TGATGAGAAGAGCGATGATGAAGAGATGATGAAGAGATGATGAAGAGAGATGATGAAG 683
466 CTGGCAGCTTCCGACGAGCGCTATCACTACAGATGATG 503
684 TGACGATGATGATGAAGAGATGATGAAGATGATGAAG 721

RESULT 36
LOCUS CNS008CA 945 bp DNA linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TERT end of BAC # BACR16N06 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL051492
VERSION AL051492.1 GI:4933546
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 945)
AUTHORS Genoscope.
TITLE Direct Submision
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqefgenoscope.cns.fr - web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end and sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammose in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES

source

Location/Qualifiers
1..945
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR16N06"
/clone_11b="RPCI-98"
/note="end : TERT3"

Query Match 2.2%; Score 52.2; DB 29; Length 945;
Best Local Similarity 39.2%; Pred. No. 1.3;
Matches 123; Conservative 36; Mismatches 155; Indels 0; Gaps 0;

1225 TAAATATTTGAAAGCAAGTATCAAAACAAGAGAGTCTTACACACTTTACTGCTTA 1284
197 TWATRAWTTTCAATWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMA 256
1285 AAAAGAAATGTTGCTCGTACCAAGAAATTTATGATTAAGCATTAATCTGTTAAC 1344
257 AAAAAMGAAAWMTTGCAGAAAAAAGAAAAAAGAAAAAATTAAMWMAWMAWMAWMA 316
1345 TGAGGCTATTAAGCCTTGTGTAAGAAATTAAGGTCGTAATTTCTGATTTCCAGCCTTAGA 1404
317 WCTTAMWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMAWMA 376
1405 CAATATTTAGACGCTTGAATGATGATCGATATATTAAGAAAAATTTGATGATGATTT 1464
377 AAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGATTT 436
1465 ATTGCATTTCTAGACCAATTAACCATCCAGAGGAGCTTGGCAACCAAAATTCCTCAAT 1524
437 AAACCAAMCAATTTGGAATTAATTTCAAAAAAAGAAAAAAGAAAAAAGAAAAAAGCAAT 496
1525 TGAGTATCTGAAG 1538
497 TAAATACAAABAR 510

RESULT 37
LOCUS BM169710 435 bp mRNA linear EST 04-DEC-2001
DEFINITION EST572233 PyBS Plasmodium yoelii yoelii cDNA clone pICPv42 5' end, mRNA sequence.
ACCESSION BM169710
VERSION BM169710.1 GI:17302942
KEYWORDS EST.
SOURCE Plasmodium yoelii yoelii
ORGANISM Plasmodium yoelii yoelii
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
REFERENCE 1 (bases 1 to 435)
AUTHORS Carlton, J.M., Daly, T.M., Long, C.A., Bergman, L.W., Valdivia, A.B., Fraser, C.M. and Carucci, D.J.
TITLE Plasmodium yoelii EST project at TIGR
JOURNAL Unpublished
COMMENT Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference Reagent Resource Center, ATCC
<http://www.malaria.mr4.org/mr4pages/index.html>
Seq primer: ADP.

FEATURES

source

Location/Qualifiers
1..435
/organism="Plasmodium yoelii yoelii"
/mol_type="mRNA"
/strain="17XL"
/db_xref="taxon:73239"

LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
CNS023KH/C	CNS023KH	853 bp	DNA	linear	GSS 01-SEP-2000	
	Tetradodon nigroviridis genome survey sequence PUC-Or1 end of clone 23JA11 of library G from Tetradodon nigroviridis, genomic survey sequence.					
		AL179594	GI:7817651			
		GSS, genome survey sequence.				
		Tetradodon nigroviridis				
		Euryarchaea, Metazoa, Chordata, Craniata; Vertebrata; Euteleostomi; Acanthopterygii; Neopterygii; Telostei; Euteleostei; Neoteleostei; Acanthomorphi; Acanthopterygii; Percormorphi; Tetraodontiformes; Tetraodontidae; Tetraodontidae; Tetradodon.				
REFERENCE	1	Roest Crolius,H., Jailon,O., Dasilva,C., Bouneau,L., Fisher,C., Bernot,A., Filames,C., Winker,P., Brotlier,P., Quetier,F., Saurin,W. and Weissenbach,J.				
TITLE	Estimate of human gene number provided by genome-wide analysis using Tetradodon nigroviridis DNA sequence					
JOURNAL	Nat. Genet. 25 (2), 235-238 (2000)					
MEDLINE	20296633					
PUBMED	10835645					
REFERENCE	2	Roest Crolius,H., Jailon,O., Dasilva,C., Ozouf-Costaz,C., Filames,C., Fischer,C., Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and Weissenbach,J.				
TITLE	Characterization and repeat analysis of the compact genome of the freshwater pufferfish Tetradodon nigroviridis					
JOURNAL	Genome Res. 10 (7), 939-949 (2000)					
MEDLINE	20359837					
PUBMED	10899143					
REFERENCE	3 (bases 1 to 853)					
TITLE	Genoscope.					
JOURNAL	Submitted (12-APR-2000) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr)					
TITLE	- Web : www.genoscope.cns.fr					
COMMENT	This sequence is a single read and was generated as part of a large scale clone-and-sequencing project of the Tetradodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetradodon .					
FEATURES	Location/Qualifiers					
source	1..853					
	/organism="Tetradodon nigroviridis"					
	/mol_type="genomic DNA"					
	/db_xref="taxon:99883"					
	/clone="23JA11"					
	/clone_11b="G"					
	/note="Genoscope sequence ID : COAG23JA06SP1-end : PUC-Or1"					
BASE COUNT	251 a 256 c 124 g 203 t 19 others					
ORIGIN						
Query Match	2.2% ; Score 51.6 ; DB 29 ; Length 853 ;					
Best Local Similarity	49.6% ; Pred. No. 1.6 ;					
Matches 126 ;	Conservative 2 ; Mismatches 126 ; Indels 0 ; Gaps 0 ;					
Oy	186 ATGGCGACCATCATTCATTATTCACATGTGAAGTTCTTATGACGCTATCATCTAGTGAAG	245				
Db	381 ATGGAGATGATGAT	322				
Oy	246 AATTACTATGAAAGATCCAAACTATTAAGTAAAGATGAGAGATATTGTAATGAGGCA	305				
Db	321 AAGATATCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	262				
Oy	306 AGGCTGATATGTTATCAAGGTAGATGGAATAACTATATGTTTACCTTAAGGATGCTGCC	365				
Db	261 ATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	202				
Oy	366 ACGCGGATACGCTCCGTCACAAAAGGAAATCATCGACAAAACAGAGCATATGTCAC	425				

Db	201	ACGTTAATATGCGATGCGATGATCATATTAACAATGCGCATGCAATGATATATATATA	142
Qy	426	ATCGTGAAGGTGCA	439
Db	141	TTGATGGTAATGAA	128
RESULT 40			
AU088119			
LOCUS			
DEFINITION			
ACCESSION	AU088119	500 bp	mRNA
VERSION	AU088119.1	GI:12390260	EST
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS	Watanabe,J., Sasaki,M., Suzuki,Y. and Sugano,S.		
TITLE	Full-malaria: a database for a full-length enriched cDNA library from human malaria parasite, Plasmodium falciparum		
JOURNAL	Nucleic Acids Res.	29 (1), 70-71	(2001)
MEDLINE	20574754		
PUBMED	11125052		
COMMENT			
	Contact: Junichi Watanabe		
	Institute of Medical Science		
	The University of Tokyo, Department of Parasitology		
	4-6-1, Shirokanebashi, Minato-ku, Tokyo 108-8639, Japan		
	Tel: 81-3-5449-5378		
	Fax: 81-3-5449-5410		
	Email: jwatanabe@manage.ims.u-tokyo.ac.jp		
	Suzuki,Y., Yoshitomo-Nakagawa,K., Maruyama,K., Suyama,A. and Sugano,S.		
	Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library		
	Gene 200 (1-2), 149-156 (1997).		
FEATURES			
source			
	1..500		
	/organism="Plasmodium falciparum 3D7"		
	/mol_type="mRNA"		
	/isolate="3D7"		
	/db_xref="taxon:36329"		
	/clone="XFpn6549"		
	/dev_stage="erythrocytic stage"		
	/clone_lib="Sugano Malaria cDNA library"		
BASE COUNT	214 a	37 c	86 g
ORIGIN			
	163 t		
Query Match	2.1%;	Score 51;	DB 9;
Best Local Similarity	50.2%;	Pred. No. 1.9;	
Matches 126;	Conservative 0;	Mismatches 125;	Indels 0;
Gaps 0;			
Qy	173	TATGTCCTCCACATGGCGACCACTATCATTTATTAACAATGGTAAGTCTCTTATGACGT	232
Db	23	TTTGATATATATGATGATTAATGATGATGAATGAATGATATATATGATGATGATGAT	82
Qy	233	ATCATCAGTGAAGATATTACTCATGTAAGAATCCAAACTATATAGCTTAAAGATGAGATATT	292
Db	83	AATATGATGATATATATATATGATTAATGACAAATATGATGATGATGATGATGATGAT	142
Qy	293	GTTATATAGGTCAAGGGTGAATATGTTATATCAAGGTAGATGGAATAATCTATGTTTACCTT	352
Db	143	AATGATGATGACAAATATATGATGATGACAAATTAATGATGATGATGATGATGATGAT	202
Qy	353	AAGGATCTCCGCCACGGCGGATTAACGTCGCCGTAACAAGAGGAATCAATGCACAAAAACA	412
Db	203	AATGATATATACAAATATGATGATGATGACAAATTAATGATGATGATGATGATGATGAT	262
Qy	413	GAGCATAGTCA	423
Db	263	GATGATATATGA	273
RESULT 41			

A2196050/c 726 bp DNA linear GSS 31-AUG-2000
 LOCUS
 DEFINITION SP 1031 A2 E05 SP6E Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library Strongylocentrotus purpuratus genomic clone Plate=1031 Col=10 Row=1, genomic survey sequence.
 ACCESSION A2196050
 VERSION A2196050.1 GI:8389873
 KEYWORDS GSS.
 SOURCE Strongylocentrotus purpuratus
 ORGANISM Strongylocentrotus purpuratus
 Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa; Echinoidae; Euechinoidae; Echinacea; Echinoida; Strongylocentrotidae; Strongylocentrotus.
 REFERENCE 1 (bases 1 to 726)
 Cameron,R.A., Mahairas,G., Raat,J.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Poustka,A.J., Livingston,B.T., Wray,G.A., Ettensohn,C.A., Lehnach,H., Britten,R.J., Davidson,E.H. and Hood,L.
 A sea urchin genome project: Sequence scan, virtual map, and additional resources
 Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
 TITLE A sea urchin genome project: Sequence scan, virtual map, and additional resources
 JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)
 MEDLINE 20402566
 PUBMED 10920195
 COMMENT Contact: Cameron, RA, Davidson, EH, Hood, L
 Division of Biology 156-29
 California Institute of Technology
 Pasadena California 91125, USA
 Tel: (626) 395-8421
 Fax: (626) 793-3047
 Email: acameron@caltech.edu
 Plate: 1031 row: 1 column: 10
 Seq primer: SP6
 Class: BAC ends
 High quality sequence stop: 726.
 Location/Qualifiers
 1..726
 /organism="Strongylocentrotus purpuratus"
 /mol_type="genomic DNA"
 /db_xref="taxon:7668"
 /clone="Plate=1031 Col=10 Row=1"
 /clone_lib="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library"
 /note="Organ: sperm; Vector: BAC3.6; BAC Clones in E-Coli DH10B"
 BASE COUNT 253 a 210 c 64 g 199 t
 ORIGIN
 Query Match 2.1%; Score 50.8; DB 28; Length 726;
 Best Local Similarity 46.8%; Pred. No. 2.3;
 Matches 160; Conservative 0; Mismatches 182; Indels 0; Gaps 0;
 Oy 256 GAAAGATCCAACTTAAGCTTAAGAGATTTGTTATGAGCGCAAGGGTGATA 315
 Db 577 GAAAGATGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 518
 Oy 316 TGTATCAAGGTAGATGAAATATCTATGTTAAGTGTGCGCCAGCGGATAA 375
 Db 517 TGAATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGA 458
 Oy 376 CGTCGTCACAAAGAGAAATCAATGACAAACAAAGACATAGTCAACATCGTGAAG 435
 Db 457 GGATGATGAGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGA 398
 Oy 436 TGGAACTCCAAAGAACGATGCTGCTTGGACAGTTGCGAAGACGTTATATAC 495
 Db 397 TGAATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 338
 Oy 496 AGATGATGATATATCTTAATGCTTGTGATATCATAGAGGATAGGATGCTTATAT 555
 Db 337 TGAATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGA 278
 Oy 556 CGTTCCTCATGAGATCATTAACATTAATCTTAAGATGA 597

Db 277 TGTGATGACGCGTATGATGATGATGATGATGATGATGATGATGATGATGATGA 236
 RESULT 42
 LOCUS CNS06X95
 DEFINITION T3 end of clone AX0AA039F08 of library AX0AA from strain CBS 7064
 ACCESSION AL149462
 VERSION AL149462.1 GI:12202640
 KEYWORDS GSS.
 SOURCE Pichia farinosa
 ORGANISM Pichia farinosa
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Pichia.
 REFERENCE 1 (bases 1 to 1007)
 Souciet,U.L., Aigle,M., Attiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,B., Brothier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Maupertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
 Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
 FEBS Lett. 487 (1), 3-12 (2000)
 TITLE Yeast species for molecular evolution studies
 JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
 MEDLINE 20584711
 PUBMED 11152876
 REFERENCE 2 (bases 1 to 1007)
 de Montigny,J., Spehner,C., Souciet,J., Tekala,F., Dujon,B., Wincker,P., Attiguenave,F. and Potier,S.
 Genomic exploration of the hemiascomycetous yeasts: 15. Pichia sorbitophila
 FEBS Lett. 487 (1), 87-90 (2000)
 JOURNAL 20584725
 MEDLINE 11152890
 PUBMED 3 (bases 1 to 1007)
 REFERENCE Genoscope.
 TITLE Direct Submission
 JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail: seqref@genoscope.cns.fr - Web: www.genoscope.cns.fr)
 COMMENT This GSS is part of a random genome sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
 Location/Qualifiers
 1..1007
 /organism="Pichia farinosa"
 /mol_type="genomic DNA"
 /strain="CBS 7064"
 /db_xref="taxon:4920"
 /clone="AX0AA039F08"
 /clone_id="AX0AA"
 /note="end: T3"
 BASE COUNT 533 a 86 c 126 g 161 t 101 others
 ORIGIN
 Query Match 2.1%; Score 50.2; DB 29; Length 1007;
 Best Local Similarity 36.5%; Pred. No. 3.5;
 Matches 130; Conservative 46; Mismatches 180; Indels 0; Gaps 0;
 Oy 1093 ACCAAATCTTAATTAAGTCAATCTCTTGTGATGATGATGATGATGATGATG 1152
 Db 29 ACCCAAGAAAGAAAGTCCGCTGTTGATGCTGTTGTTAATTAAGAAAGAAAG 88
 Oy 1153 GAAAGATATGATATTCAGAAAGGAGCATCTCGTTATGCTTTGCGAAAGATTACC 1212

Db 89 MAAGGMAAAMAAWAAAAAAMAAAGCAATCGTCTCTGAGGAAAAAA 148

Qy 1213 ATCTGAACCTGTTAAATCTTGAAGCAAGTATCAAAACAGAGNGTTCACACAC 1272

Db 149 AATATMAAAAMWTTAAAMMAAAATATTAAGTTAAAAAGMAAGCAATCAAAAT 208

Qy 1273 TTTACTGCTAAAAAAGAAATGTCTCTCTGTCGACCAAGATTTATGATTAACATA 1332

Db 209 AAAAAAMMAAAACAMGTCATGTGCMGSGTGATGAMGTBAWTATWAGMAAAABAMR 268

Qy 1333 TTAATCTGTTAAGAGGCTCATTAAGCTTTGTTAAATAGGCTGCTAATTCGATTT 1392

Db 269 TAAAGWMTAAAGMAAAMAAAMAAAGAGGTTTAAAMAAAMAAAMAAAMAA 328

Qy 1393 CCAAGCCTTAGACAAATTTATAGACGCTTGAATGATGATGACCTAATTAAGAAA 1448

Db 329 AAAAAAAGCAAAAAAAMAAAMAAAMAAAMAAAMAAAMAAACGAAMWAA 384

RESULT 43

LOCUS BM167469 445 bp mRNA linear EST 04-DEC-2001

DEFINITION EST569992 PyBS Plasmodium yoelii yoelii cDNA clone PYCOK56 5' end, mRNA sequence.

ACCESSION BM167469

VERSION BM167469.1 GI:17300701

KEYWORDS EST.

SOURCE Plasmodium yoelii yoelii

ORGANISM Plasmodium yoelii yoelii

REFERENCE 1 (bases 1 to 445)

AUTHORS Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Valdiva,A.B., Fraser,C.M. and Carucci,D.J.

TITLE Plasmodium yoelii EST project at TIGR

JOURNAL Unpublished

COMMENT Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-858-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADP.

FEATURES

source

Location/Qualifiers

1..445

/organism="Plasmodium yoelii yoelii"

/mol_type="mRNA"

/strain="17X1"

/db_xref="taxon:73239"

/clone="PYCOK56"

/dev_stage="Asexual blood stages"

/lab_host="E. coli XL-1 Blue"

/clone_lib="PyBS"

/note="Vector: PAD-GAL4; At 20-25% parasitemia, blood was collected from BALB/cByJ mice infected with Py17XL parasites, and leukocytes removed by passage over microcrystalline cellulose columns. Total RNA was isolated using the guanidium isothiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase, and EcoRI adaptors ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybriZAP arms directionally using EcoRI-XhoI cleaved arms. After packaging, the phagemid vector (PAD-GAL4) was excised from the HybriZAP vector and plasmid DNA isolated."

BASE COUNT 173 a 36 c 145 g 91 t

ORIGIN

Query Match 2.1%; Score 50; DB 12; Length 445;

Best Local Similarity 46.7%; Pred. No. 2.9;

Matches 158; Conservative 0; Mismatches 180; Indels 0; Gaps 0;

Qy 260 GATCCAACTATATAGCTAAAGATAGATATTTGTAATGAGTCAAGGCTGATGTT 319

Db 61 GACACAGGAAATAGACAGCAAGAGGCGATAGAGATGATGAAGATGAAGATGAT 120

Qy 320 ATCAAGTAGATGAAAAATCTATGTTTACCTTAAGATGCTGCCACCGGATAAGTC 379

Db 121 GATGACGACGATGAAGATGACGATGAAGCGATGATGATGAAGACGACGATGAT 180

Qy 380 CCTACAAAAGAGGAATCATGACAAAAACAAGACATAGTCACATGTTGAAGTGA 439

Db 181 GATGACGATGATGACGATGACGATGATGATGATGATGATGATGATGAAGAGAT 240

Qy 440 ACTCCAGAAACGATGCTGCTGCTGCTGCGACGCTGCGACGACGCTATACAGAT 499

Db 241 GACGATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300

Qy 500 GATGCTTATATCTTAAATGCTTCTGATATCATAGAGATATGCTGATGCTTATGCT 559

Db 301 GATGATGATGACGACGATGAAGATTTTGAAGATGATGATGATGATGATGATGATG 360

Qy 560 CCTCATGAGATCATTCATTCATTCATTCCTTAAGATGA 597

Db 361 GACGATGATGACGATGATGAAGACGATGATGATGATGA 398

RESULT 44

LOCUS BU496963 555 bp mRNA linear EST 11-SEP-2002

DEFINITION PFESToab61c09.y1 Plasmodium falciparum 3D7 asexual cDNA Plasmodium falciparum 3D7 cDNA 5', mRNA sequence.

ACCESSION BU496963

VERSION BU496963.1 GI:22793167

KEYWORDS EST.

SOURCE Plasmodium falciparum 3D7

ORGANISM Plasmodium falciparum 3D7

REFERENCE 1 (bases 1 to 555)

AUTHORS Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D., Marra,M., Hillier,L., Martin,J., Wylie,T., Dame,W., Theisling,B., Bowers,Y., Gibbons,M., Ritey,E., Bennett,J., Jentes,E., Ronko,I., Tsagarisshvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A., Maguire,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D., Waterston,R., Wilson,R. and Sibley,D.

TITLE WashU Plasmodium EST Project

JOURNAL Unpublished

COMMENT Contact: L. David Sibley
WashU Plasmodium EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.wustl.edu
Library was constructed by Depopam Chakrabarti DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: L. David Sibley (sibley@orcim.wustl.edu), Washington University
Seq primer: -40UP from Gibco
High quality sequence stop: 421.

FEATURES

source

Location/Qualifiers

1..555

/organism="Plasmodium falciparum 3D7"

/mol_type="mRNA"

/db_xref="taxon:36329"

/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"

/clone_lib="Plasmodium falciparum 3D7 asexual cDNA"

/note="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:

[illegible]

```

TITLE
JOURNAL
COMMENT
Plasmodium yoelli EST project at TIGR
Unpublished
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mrt.org/mr4pages/index.html
Seq primer: AdF

FEATURES
Source
Location/Qualifiers
1..605
/organism="Plasmodium yoelli yoelli"
/mol_type="mRNA"
/strain="17XL"
/db_xref="taxon:73239"
/clone="PYLG22"
/dev_stage="asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/clone_lib="PyBS"
/notes="Vector: PAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/cBY mice infected with PY17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was

```

isolated using the guanidium thiocyanate method, and mRNA isolated using oligo(dT)-cellulose chromatography. First strand cDNA synthesis was completed using a 50-base primer and reverse transcriptase in the presence of 5-methyl dCTP. After second strand synthesis, uneven termini were treated with Pfu DNA polymerase and EcoRI

adducts ligated to the blunt ends. The sample was cleaved with XhoI and separated on a Sephacryl S-500 column. Size-fractionated cDNA was precipitated and ligated to HybridZap arms directionally using EcoRI. XhoI cleaved arms. After packaging, the phagemid vector (pMD-GM4) was excised from the HybridZap vector and plasmid DNA

[illegible][illegible]

RESULT	48
LOCUS	BJ073162
DEFINITION	BJ073162 NIBB Mochii normalized Xenopus tailbud library Xenopus laevis cDNA clone XU109D17 5', mRNA sequence.
ACCESSION	BJ073162
VERSION	BJ073162.1 GI:17503351
KEYWORDS	EST.
SOURCE	Xenopus laevis (African clawed frog)
ORGANISM	Xenopus laevis
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipridae; Xenopodinae; Xenopus.
AUTHORS	1 (bases 1 to 629) Kitayama,A., Terasaka,C., Mochii,M., Ueno,N., Shin-i,T. and Kohara,Y.
TITLE	Expressed genes in X. laevis embryo
JOURNAL	Unpublished
COMMENT	Contact: Tadasu Shin-i Center For Genetic Resource Information National Institute of Genetics 1111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tshini@genes.nig.ac.jp. Location/Qualifiers 1..629 /organism="Xenopus laevis" /mol_type="mRNA" /db_xref="taxon:8355" /clone="XU109D17" /tissue_type="whole embryo" /dev_stage="stage 25" /clone_lib="NIBB Mochii normalized Xenopus tailbud library"
FEATURES	
source	
BASE COUNT	207 a 109 c 165 g 148 t
ORIGIN	
Query Match	2.1%; Score 49.4; DB 12; Length 629;
Best Local Similarity	49.8%; Pred No. 4.5;

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 687)
Eukaryote: Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
Carlton,J.M., Daly,T.M., Long,C.A., Bergman,L.W., Vaidya,A.B.,
Fraser,C.M. and Carucci,D.J.
Plasmodium yoelii EST project at TIGR
Unpublished
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
http://www.malaria.mr4.org/mr4pages/index.html
Seq primer: ADP.

FEATURES
Source
Location/Qualifiers
1..687
/organism="Plasmodium yoelii yoelii"
/mol_type="mRNA"
/strain="17XL"
/db_xref="taxon:73239"
/clone="PYCPU72"
/dev_stage="Asexual blood stages"
/lab_host="E. coli XL-1 Blue"
/clone_lib="PYBS"
/note="Vector: pAD-GAL4; At 20-25% parasitemia, blood was
collected from BALB/cBYJ mice infected with Py17XL
parasites, and leukocytes removed by passage over
microcrystalline cellulose columns. Total RNA was
isolated using the guanidium isothiocyanate method, and
mRNA isolated using oligo(dT)-cellulose chromatography.
First strand cDNA synthesis was completed using a 50-base
primer and reverse transcriptase in the presence of
5-methyl dCTP. After second strand synthesis, uneven
termini were treated with Pfu DNA polymerase and EcoRI
adaptors ligated to the blunt ends. The sample was cleaved
with XhoI and separated on a Sephacryl S-500 column.
Size-fractionated cDNA was precipitated and ligated to

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comphen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 13, 2003, 21:04:03 ; Search time 147 Seconds
(without alignments)
7173.227 Million cell updates/sec

Title: US-09-765-271-55

Perfect score: 2389
Sequence: 1 TTCTTACGAGTGGAGCTGT.....TAAGTACGAAAAATAAAC 2389

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Issued Patents NA:*

1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*\n2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*\n3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*\n4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*\n5: /cgn2_6/prodata/2/ina/PCUTUS_COMB.seq:*\n6: /cgn2_6/prodata/2/ina/backfillseq1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2388	100.0	2389	3	US-08-961-083-55
2	2388	100.0	2389	4	US-09-536-784-55
3	2388	100.0	2451	4	US-09-468-656A-9
4	2388	100.0	8195	4	US-08-961-527-94
5	991	41.5	2531	4	US-09-468-656A-11
6	990	41.4	2531	4	US-09-468-656A-5
7	987.6	41.3	2290	3	US-08-961-083-65
8	987.6	41.3	2290	4	US-09-536-784-65
9	653.8	27.4	2359	4	US-08-961-527-243
10	385.4	16.1	1342	3	US-08-961-083-181
11	385.4	16.1	1342	4	US-09-536-784-181
12	385.4	16.1	1455	4	US-09-468-656A-7
13	385.4	16.1	6867	4	US-08-961-527-192
14	381.4	16.0	973	4	US-08-961-527-355
15	243.4	10.2	1684	4	US-08-961-527-258
16	109	4.6	841	2	US-08-743-637B-34
17	109	4.6	841	3	US-08-526-840B-34
18	57.4	2.4	7218	1	US-08-232-463-14
19	51.6	2.2	11091	4	US-09-134-001C-2243
20	45.2	1.9	2277	1	US-08-676-974-2
21	45.2	1.9	2277	2	US-08-676-974-2
22	45.2	1.9	2277	2	US-09-098-487-2
23	41.4	1.7	5361	3	US-08-973-462-2
24	41.4	1.7	6152	3	US-08-973-462-1
25	41.2	1.7	696	3	US-09-461-697-193
26	41.2	1.7	699	3	US-09-461-697-191
27	41.2	1.7	717	3	US-09-461-697-189

28	41.2	1.7	774	3	US-09-461-697-187	Sequence 187, App
29	41.2	1.7	819	3	US-09-461-697-185	Sequence 185, App
30	41.2	1.7	1280	4	US-09-220-133-155	Sequence 155, App
31	41.2	1.7	1669	3	US-09-461-697-184	Sequence 184, App
32	41	1.7	5340	4	US-09-627-122-21	Sequence 21, Appl
33	41	1.7	14066	4	US-09-601-198-56	Sequence 56, Appl
34	40.6	1.7	1704	4	US-09-134-001C-925	Sequence 925, App
35	40.2	1.7	792	4	US-09-328-352-2472	Sequence 2472, App
36	40	1.7	30349	4	US-09-134-001C-1027	Sequence 1027, App
37	39.8	1.7	297	4	US-09-134-001C-1027	Sequence 1027, App
38	39.8	1.7	1116	4	US-09-134-001C-1028	Sequence 1028, App
39	39.4	1.7	1308	4	US-09-601-198-3	Sequence 3, Appl
40	39.4	1.6	1664976	4	US-08-916-421B-1	Sequence 1, Appl
41	39.2	1.6	929	4	US-09-671-317-14	Sequence 14, Appl
42	39	1.6	1839	4	US-09-134-001C-1191	Sequence 1191, App
43	38.8	1.6	1856	2	US-08-360-606B-29	Sequence 29, Appl
44	38.8	1.6	2150	2	US-08-861-464-13	Sequence 13, Appl
45	38.8	1.6	2150	2	US-08-396-001-13	Sequence 13, Appl
46	38.6	1.6	2150	3	US-09-323-433A-13	Sequence 13, Appl
47	38.6	1.6	6744	1	US-08-119-125A-2	Sequence 2, Appl
48	38.6	1.6	6744	1	US-08-446-855A-1	Sequence 1, Appl
49	38.6	1.6	8920	2	US-09-150-741-1	Sequence 1, Appl
50	38.4	1.6	5319	1	US-08-169-927-1	Sequence 1, Appl
51	38.4	1.6	6846	4	US-08-961-527-198	Sequence 198, App
52	38.4	1.6	580073	4	US-08-545-528D-1	Sequence 1, Appl
53	38.2	1.6	1866	4	US-09-601-198-153	Sequence 153, App
54	38	1.6	1001	4	US-09-671-317-439	Sequence 439, App
55	38	1.6	3280	1	US-08-259-000-4	Sequence 4, Appl
56	38	1.6	3280	1	US-08-729-767-6	Sequence 6, Appl
57	37.8	1.6	1664976	4	US-08-916-421B-1	Sequence 1, Appl
58	37.6	1.6	438	4	US-09-134-001C-1064	Sequence 1064, App
59	37.6	1.6	2000	4	US-09-632-313A-1	Sequence 1, Appl
60	37.6	1.6	2000	4	US-09-921-942B-6	Sequence 6, Appl
61	37.4	1.6	3294	4	US-08-923-924A-7	Sequence 7, Appl
62	37.4	1.6	26385	4	US-08-961-527-3	Sequence 3, Appl
63	37.2	1.6	462	4	US-09-328-352-2077	Sequence 2077, App
64	37.2	1.6	1079	1	US-07-781-355-1	Sequence 1, Appl
65	37.2	1.6	2397	4	US-09-134-001C-1080	Sequence 2080, App
66	37.2	1.6	1830121	4	US-09-557-884-1	Sequence 1, Appl
67	37.2	1.6	1830121	4	US-09-643-990A-1	Sequence 1, Appl
68	37	1.5	1103	3	US-09-122-400B-20	Sequence 20, Appl
69	37	1.5	1167	3	US-08-592-658-1	Sequence 1, Appl
70	37	1.5	1893	4	US-09-328-352-112	Sequence 112, App
71	36.6	1.5	282	3	US-09-461-697-205	Sequence 205, App
72	36.6	1.5	306	3	US-09-461-697-203	Sequence 203, App
73	36.6	1.5	807	4	US-09-328-352-2911	Sequence 2911, App
74	36.6	1.5	1065	4	US-09-107-532A-3579	Sequence 48, Appl
75	36.6	1.5	2287	4	US-09-059-584-47	Sequence 47, Appl
76	36.6	1.5	2287	4	US-09-059-584-47	Sequence 58, Appl
77	36.6	1.5	3337	1	US-08-601-198-58	Sequence 1, Appl
78	36.6	1.5	3337	1	US-08-072-610-1	Sequence 1, Appl
79	36.6	1.5	3337	2	US-08-719-822B-1	Sequence 1, Appl
80	36.6	1.5	3337	3	US-09-092-458-1	Sequence 1, Appl
81	36.6	1.5	3387	1	US-08-468-557-1	Sequence 1, Appl
82	36.6	1.5	4766	5	PCR-US93-07261-10	Sequence 10, Appl
83	36.4	1.5	1891	3	US-08-973-462-3	Sequence 3, Appl
84	36.2	1.5	231	3	US-09-461-697-207	Sequence 207, App
85	36.2	1.5	2201	4	US-09-642-000-1	Sequence 1, Appl
86	36.2	1.5	2223	4	US-08-257-073-4	Sequence 4, Appl
87	36.2	1.5	3095	6	5231168-1	Sequence 5, Appl
88	36.2	1.5	6008	1	US-07-789-915A-5	Sequence 5, Appl
89	36.2	1.5	6008	1	US-08-005-002C-5	Sequence 5, Appl
90	36.2	1.5	6008	1	US-08-487-202A-5	Sequence 5, Appl
91	36	1.5	1151	1	US-08-236-754-3	Sequence 3, Appl
92	36	1.5	3365	4	US-09-134-001C-1983	Sequence 1983, App
93	36	1.5	3365	4	US-09-134-001C-2311	Sequence 2311, App
94	36	1.5	5565	3	US-09-068-195-21	Sequence 21, Appl
95	35.8	1.5	567	4	US-07-835-412B-2	Sequence 2, Appl
96	35.8	1.5	951	2	US-08-308-887A-2	Sequence 2, Appl
97	35.8	1.5	951	2	US-08-881-094-2	Sequence 224, App
98	35.8	1.5	1886	4	US-08-936-165A-224	Sequence 668, App
99	35.8	1.5	2183	4	US-09-620-312D-3578	
100	35.8	1.5	2183	4	US-09-620-312D-3578	

ALIGNMENTS

RESULT 1

US-08-961-083-55

Sequence 55, Application US/08961083

Patent No. 6159469

GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,083

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ. ID NO: 55:

SEQUENCE CHARACTERISTICS:

LENGTH: 2389 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

US-08-961-083-55

Query Match 100.0%; Score 2389; DB 3; Length 2389;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 TTCTTACGAGTGGAGCTGTATCAAGCTAAGAAAGGTTAAAGAAATATCGTTTCTTA 60
DB 1 TTCTTACGAGTGGAGCTGTATCAAGCTAAGAAAGGTTAAAGAAATATCGTTTCTTA 60
OY 61 TATAGATGAAAAACAAGCGAGCGCAAAAAAGGAGAAATTTGACTCTCTGATGAGGTTAGCA 120
DB 61 TATAGATGAAAAACAAGCGAGCGCAAAAAAGGAGAAATTTGACTCTCTGATGAGGTTAGCA 120
OY 121 GCGTGAAGAAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTAC 180
DB 121 GCGTGAAGAAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTAC 180
OY 121 GCGTGAAGAAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTAC 180
DB 121 GCGTGAAGAAATCAATGCTGAGCAAAATCGTCATCAAGATAACAGACCAAGGCTATGTAC 180
OY 181 TTTCATGCGAGCACTATCATTTATTAAGTGAAGGTTCTTATGACGCTATATATG 240
DB 181 TTTCATGCGAGCACTATCATTTATTAAGTGAAGGTTCTTATGACGCTATATATG 240
OY 181 TTTCATGCGAGCACTATCATTTATTAAGTGAAGGTTCTTATGACGCTATATATG 240
DB 181 TTTCATGCGAGCACTATCATTTATTAAGTGAAGGTTCTTATGACGCTATATATG 240
OY 241 TGAAGAAATTAATCAATGAGTCAAACTATTAAGTGAAGTGAAGTGAAGTGAAGTGAAG 300
DB 241 TGAAGAAATTAATCAATGAGTCAAACTATTAAGTGAAGTGAAGTGAAGTGAAGTGAAG 300
OY 301 GGTCAAGGTTGATATGTTATCAAGGTAGATGAAATTAATGTTTACCTTAAGATGC 360
DB 301 GGTCAAGGTTGATATGTTATCAAGGTAGATGAAATTAATGTTTACCTTAAGATGC 360
```

```
DB 301 GGTCAAGGTTGATATGTTATCAAGGTAGATGAAATTAATGTTTACCTTAAGATGC 360
OY 361 TGCCACGCGGATTAACGTCGTACAAAAAGAGAAATCAATGACAAAAACAAGACATAG 420
DB 361 TGCCACGCGGATTAACGTCGTACAAAAAGAGAAATCAATGACAAAAACAAGACATAG 420
OY 421 TCAACATGCTGAAGTGAATCCCAAGAAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
DB 421 TCAACATGCTGAAGTGAATCCCAAGAAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
OY 481 AGAAGCTATATCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
DB 481 AGAAGCTATATCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 540
OY 541 TGGGATGCTTATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
DB 541 TGGGATGCTTATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600
OY 601 ATCAGCTAGCAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
DB 601 ATCAGCTAGCAGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660
OY 661 TTCAAGAACTTATGCGGACAAATAGCGATACACTTCAAGAACTGCGTACCTTC 720
DB 661 TTCAAGAACTTATGCGGACAAATAGCGATACACTTCAAGAACTGCGTACCTTC 720
OY 721 TGTAGCAATCCAGAACTCAATTAATCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 780
DB 721 TGTAGCAATCCAGAACTCAATTAATCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCA 780
OY 781 AGCAAGTCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
DB 781 AGCAAGTCAAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 840
OY 841 TCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
DB 841 TCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
OY 901 AGCTAGAGGTTGACGATGACGACGAGGATCAATTAACATTAACATTAACATTAACAT 960
DB 901 AGCTAGAGGTTGACGATGACGACGAGGATCAATTAACATTAACATTAACATTAACAT 960
OY 961 GTCTGAATGGAGAAAGAAATCGCTGATTAATCCCTTCTGTTATGTTTCAACATTTG 1020
DB 961 GTCTGAATGGAGAAAGAAATCGCTGATTAATCCCTTCTGTTATGTTTCAACATTTG 1020
OY 1021 GGTACAGATTCAGGCGAGAAACCAAGTCCAAACCGACTCCGGAACCTTACTGACG 1080
DB 1021 GGTACAGATTCAGGCGAGAAACCAAGTCCAAACCGACTCCGGAACCTTACTGACG 1080
OY 1081 CCGGCAACCTGACCAAAATCTTAATAGACTCAAAATCTTCTGTTAGTCAAGCTGGT 1140
DB 1081 CCGGCAACCTGACCAAAATCTTAATAGACTCAAAATCTTCTGTTAGTCAAGCTGGT 1140
OY 1141 ACGAAAAGTTGGGAGGATATGATTCGAAAGAAAGGAGCTCTGTTATGTTTCTTGGC 1200
DB 1141 ACGAAAAGTTGGGAGGATATGATTCGAAAGAAAGGAGCTCTGTTATGTTTCTTGGC 1200
OY 1201 GAAAGATTTACATCTGAATCTGTTAAATCTTGAAGCAAGTTATCAAAACAAGAGAG 1260
DB 1201 GAAAGATTTACATCTGAATCTGTTAAATCTTGAAGCAAGTTATCAAAACAAGAGAG 1260
OY 1261 TGTTCACACACTTAATCTGCTAATAAAGAAAGTGGTCTGCTGCTGCTGCTGCTGCTG 1320
DB 1261 TGTTCACACACTTAATCTGCTAATAAAGAAAGTGGTCTGCTGCTGCTGCTGCTGCTG 1320
OY 1321 TGATAAGCATATATATCTGTTAACTGAGGCTCAATTAAGCTTGTGTTGAAATTAAGG 1380
DB 1321 TGATAAGCATATATATCTGTTAACTGAGGCTCAATTAAGCTTGTGTTGAAATTAAGG 1380
OY 1381 TAAATTCGATTTCCAGCCTTGAACAAATTAATTAAGACCTTGAATGATGATGATGATG 1440
DB 1381 TAAATTCGATTTCCAGCCTTGAACAAATTAATTAAGACCTTGAATGATGATGATGATG 1440
```


Db 481 AGGAGCTAFACTACAGATGATGTTATATCTTTAATGCTTCTGATATCATATGAGAGATAC 540
 Oy 541 TGGTGAATGCTTATATGCTTCCCTCAGAGATCATTAACATTAATCTCTAAGAAATGAGAT 600
 Db 541 TGGTGAATGCTTATATGCTTCCCTCAGAGATCATTAACATTAATCTCTAAGAAATGAGAT 600
 Oy 601 ATCAGCTAGCGAGTGGCTGCTGTCAGAAAGCCTTCTATCTGGTCGAGGAAATCTGTCAA 660
 Db 601 ATCAGCTAGCGAGTGGCTGCTGTCAGAAAGCCTTCTATCTGGTCGAGGAAATCTGTCAA 660
 Oy 661 TTCAAAGACTATGCGCCGACAAATAGCGATTAACCTTCAAGAACAACTGGGTACCTTC 720
 Db 661 TTCAAAGACTATGCGCCGACAAATAGCGATTAACCTTCAAGAACAACTGGGTACCTTC 720
 Oy 721 TGTAAAGCAATCCAGAACTACAAATATCTAACCAAGCAACACACAACTAATACATCA 780
 Db 721 TGTAAAGCAATCCAGAACTACAAATATCTAACCAAGCAACACACAACTAATACATCA 780
 Oy 781 AGCAAGTCAAAAGTAATGACATTAATGATGCTCTTGAAGAGCTCTCAAACTGCTTTGAG 840
 Db 781 AGCAAGTCAAAAGTAATGACATTAATGATGCTCTTGAAGAGCTCTCAAACTGCTTTGAG 840
 Oy 841 TCAAAGACATGTAGAATCTGATGCGCTTGTCTTTGATCCAGACAAATCAAGTCTGAAC 900
 Db 841 TCAAAGACATGTAGAATCTGATGCGCTTGTCTTTGATCCAGACAAATCAAGTCTGAAC 900
 Oy 901 AGCTAGAGTGTGGAGTGCACAGGAGATCATTAACATTAATCTCTCAATCTCAAT 960
 Db 901 AGCTAGAGTGTGGAGTGCACAGGAGATCATTAACATTAATCTCTCAATCTCAAT 960
 Oy 961 GCTCAATTTGGAAGAAAGCAATGCTGATTAATCTCTGCTGCTTCAACATTTG 1020
 Db 961 GCTCAATTTGGAAGAAAGCAATGCTGATTAATCTCTGCTGCTTCAACATTTG 1020
 Oy 1021 GGTACCAATTCAGAGCCAGAAACAAGTCCACACGACTCCGGAACCTAGTCCAG 1080
 Db 1021 GGTACCAATTCAGAGCCAGAAACAAGTCCACACGACTCCGGAACCTAGTCCAG 1080
 Oy 1081 CCCCAGAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTGTTAGTCACTGCT 1140
 Db 1081 CCCCAGAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTGTTAGTCACTGCT 1140
 Oy 1141 ACGAAAGTGGGGAGAGATATGTAATCGAAGAAAGGGCATCTCTGTTATGCTTTGC 1200
 Db 1141 ACGAAAGTGGGGAGAGATATGTAATCGAAGAAAGGGCATCTCTGTTATGCTTTGC 1200
 Oy 1201 GAAAGATTACCATCTGAAAATCTTAAATCTTGAAGCAAGTTATCAAAACAAGAGAG 1260
 Db 1201 GAAAGATTACCATCTGAAAATCTTAAATCTTGAAGCAAGTTATCAAAACAAGAGAG 1260
 Oy 1261 TGTTCACACACTTAACTGCTTAAAGAAAGAAATGTTGCTCTCTGTCGACCAAGATTTTA 1320
 Db 1261 TGTTCACACACTTAACTGCTTAAAGAAAGAAATGTTGCTCTCTGTCGACCAAGATTTTA 1320
 Oy 1321 TGATTAAGCATTAATCTGTTAACTGAGGCTCATTAAGGCTGTTGTAAGAAATGAGGTCG 1380
 Db 1321 TGATTAAGCATTAATCTGTTAACTGAGGCTCATTAAGGCTGTTGTAAGAAATGAGGTCG 1380
 Oy 1381 TAAATTCGATTTCCAGGCTTGAACAAATTAATTAAGAGCTTGAATGATGATCACTTA 1440
 Db 1381 TAAATTCGATTTCCAGGCTTGAACAAATTAATTAAGAGCTTGAATGATGATCACTTA 1440
 Oy 1441 TAAAGAAATTTGGTATGATGATTTATTTGGCATCTCTAGACCAATTAATTAATTAAGAG 1500
 Db 1441 TAAAGAAATTTGGTATGATGATTTATTTGGCATCTCTAGACCAATTAATTAATTAAGAG 1500
 Oy 1501 ACTTGGCAAAACCAATTTCTCAAAATGAGTATCTGAAGCAAGTTCGATTTGCTCAAT 1560
 Db 1501 ACTTGGCAAAACCAATTTCTCAAAATGAGTATCTGAAGCAAGTTCGATTTGCTCAAT 1560
 Oy 1561 AGCTGATTAAGTATCAAGCTGATGTTTATGTTTGAAGCAATGATTAATCACTGA 1620
 Db 1561 AGCTGATTAAGTATCAAGCTGATGTTTATGTTTGAAGCAATGATTAATCACTGA 1620

Db 1561 AGCTGATTAAGTATCAAGCTGATGTTTATGTTTGAAGCAATGATTAATCACTGA 1620
 Oy 1621 TGAAGAGATGCATATGTAAGAGCTCCTCATATGGGCTAGTCACTGATTTGGAAGATAG 1680
 Db 1621 TGAAGAGATGCATATGTAAGAGCTCCTCATATGGGCTAGTCACTGATTTGGAAGATAG 1680
 Oy 1681 CTTTCTGATTAAGGAAAGTTGAGCTCAAGCTTATATTAAGAAAGATATCTTACC 1740
 Db 1681 CTTTCTGATTAAGGAAAGTTGAGCTCAAGCTTATATTAAGAAAGATATCTTACC 1740
 Oy 1741 TCCATCTCCAGACCAATGTTAAAGCAATCCAACTGAGATGTCAGAGCTATTTA 1800
 Db 1741 TCCATCTCCAGACCAATGTTAAAGCAATCCAACTGAGATGTCAGAGCTATTTA 1800
 Oy 1801 CAATGCTGTGAAGGAGGAAAGCAATTTCCACTGCTTCACTTCCATATATGTTGAGCA 1860
 Db 1801 CAATGCTGTGAAGGAGGAAAGCAATTTCCACTGCTTCACTTCCATATATGTTGAGCA 1860
 Oy 1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTAATCTCTATTAAGGATCATTAATTAAT 1920
 Db 1861 TACAGTTGAGGTTAAAAACGGTAATTTGATTAATCTCTATTAAGGATCATTAATTAAT 1920
 Oy 1921 TAAATTTGCTTGGTTGATGATCAACATTAACAGTCCAAATGCTATTAATCTTGAAGA 1980
 Db 1921 TAAATTTGCTTGGTTGATGATCAACATTAACAGTCCAAATGCTATTAATCTTGAAGA 1980
 Oy 1981 TTTGTTTGCAGAGTAAAGTAACTGATGAGAAACCTGAGCAAGTCCCAATTTGAATGA 2040
 Db 1981 TTTGTTTGCAGAGTAAAGTAACTGATGAGAAACCTGAGCAAGTCCCAATTTGAATGA 2040
 Oy 2041 TGGATGGGGCAATCCAGTGCAGATGTTGTTAGGCAAGAAAGCAAGTGAAGATCCAA 2100
 Db 2041 TGGATGGGGCAATCCAGTGCAGATGTTGTTAGGCAAGAAAGCAAGTGAAGATCCAA 2100
 Oy 2101 TAAAGACTTCAAGCGGATGAAGAGCAATGAGAGAAACCTTCTGAGCCAGAAATGCC 2160
 Db 2101 TAAAGACTTCAAGCGGATGAAGAGCAATGAGAGAAACCTTCTGAGCCAGAAATGCC 2160
 Oy 2161 TCAAGTGAAGACTGAAAGAGAGAGAGGCAACTCAAGGAGCAAGATTTGCTTGGCAA 2220
 Db 2161 TCAAGTGAAGACTGAAAGAGAGAGAGGCAACTCAAGGAGCAAGATTTGCTTGGCAA 2220
 Oy 2221 AGTAACGGATTTCTAGTCTGAAAGCAATGCAACGAAATCTTACCTGTTACGAAATTA 2280
 Db 2221 AGTAACGGATTTCTAGTCTGAAAGCAATGCAACGAAATCTTACCTGTTACGAAATTA 2280
 Oy 2281 TTTGACTCTTCAAAATTTGATTAACAATGATATGATGAGCAGAGCAAGAAATTAATTC 2340
 Db 2281 TTTGACTCTTCAAAATTTGATTAACAATGATATGATGAGCAGAGCAAGAAATTAATTC 2340
 Oy 2341 GTTGTAAAGAGAGTAACTCTTCAATCTGTAAGTAAAGAAATTAATTAAC 2389
 Db 2341 GTTGTAAAGAGAGTAACTCTTCAATCTGTAAGTAAAGAAATTAATTAAC 2389

RESULT 3
 US-09-468-656A-9
 ; Sequence 9, Application US/09468656A
 ; Patent No. 6582706
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John B.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
 ; TITLE OF INVENTION: Motifs
 ; FILE REFERENCE: 469201-444
 ; CURRENT FILING DATE: 1999-12-02
 ; PRIORITY FILING DATE: 1998-12-21
 ; NUMBER OF SEQ ID NOS: 14
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 9

; LENGTH: 2451
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (1)..(2451)
 ; OTHER INFORMATION: n = a, c, t or g
 us-09-468-656A-9

Query Match 100.0%; Score 2388; DB 4; Length 2451;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGAGCTGATCAAGCTAGAACGTTAAAGAAATAATCGTCTTCTTA 60
 Db 60 TTCTTACGAGTTGGAGCTGATCAAGCTAGAACGTTAAAGAAATAATCGTCTTCTTA 119
 QY 61 TTTAGATGAAAAACAAGCCGCAAAAAACGGAATTTTGACTCTGATGAGTTAGCAA 120
 Db 120 TTTAGATGAAAAACAAGCCGCAAAAAACGGAATTTTGACTCTGATGAGTTAGCAA 179
 QY 121 GCGTAGAAGATCAATGCTGAGCAATCGTCAATCAAGATPAACAGCAAGGCTATGTCAC 180
 Db 180 GCGTAGAAGATCAATGCTGAGCAATCGTCAATCAAGATPAACAGCAAGGCTATGTCAC 239
 QY 181 TTCACATGCGCAACCACTATCATTTATTAACAATGTTAGGTTCTTATGACGCTATCATCAG 240
 Db 240 TTCACATGCGCAACCACTATCATTTATTAACAATGTTAGGTTCTTATGACGCTATCATCAG 299
 QY 241 TGAAGATTTACTCATGAAAAATGCCAACTATTAAGCTTAAAGATGAGGATATTGTTAATGA 300
 Db 300 TGAAGATTTACTCATGAAAAATGCCAACTATTAAGCTTAAAGATGAGGATATTGTTAATGA 359
 QY 301 GGTCAAGGATGATATGTTATCAAGTATGAGTAAATTAATGTTTACTTAAAGATGC 360
 Db 360 GGTCAAGGATGATATGTTATCAAGTATGAGTAAATTAATGTTTACTTAAAGATGC 419
 QY 361 TSCCCACGCGGATTAAGCTCCGTAACAAGAGGAAATCAATCGCAAAAAACAAGAGCATAG 420
 Db 420 TSCCCACGCGGATTAAGCTCCGTAACAAGAGGAAATCAATCGCAAAAAACAAGAGCATAG 479
 QY 421 TCAACATCGTGAAGGTGAACTTCAAGAAACGATGCTGCTGTTGCTTGGACAGTTCCGA 480
 Db 480 TCAACATCGTGAAGGTGAACTTCAAGAAACGATGCTGCTGTTGCTTGGACAGTTCCGA 539
 QY 481 AAGAGCTTACTTACAGATGATGCTTATATCTTTAATGCTTGCATATCATAGAGATAC 540
 Db 540 AAGAGCTTACTTACAGATGATGCTTATATCTTTAATGCTTGCATATCATAGAGATAC 599
 QY 541 TGGTATGCTTATATGCTTCTCATGAGATCATTAACATTAATCTTAAGATGATG 600
 Db 600 TGGTATGCTTATATGCTTCTCATGAGATCATTAACATTAATCTTAAGATGATG 659
 QY 601 ATCAGCTAGCGAGTTGGCTGCTGCAGAGCCTTCTATCTGTGAGGAATCTGTCAAA 660
 Db 660 ATCAGCTAGCGAGTTGGCTGCTGCAGAGCCTTCTATCTGTGAGGAATCTGTCAAA 719
 QY 661 TTCAAGAACCTTATCGCCGCAAAAATGCGATTAACATTCAGAGAACCTGGGTTACCTTC 720
 Db 720 TTCAAGAACCTTATCGCCGCAAAAATGCGATTAACATTCAGAGAACCTGGGTTACCTTC 779
 QY 721 TGTAAAGCATCCAGAACTACAAATCTTAACACCAAGCAACACAGAACACTTAACGTGA 780
 Db 780 TGTAAAGCATCCAGAACTACAAATCTTAACACCAAGCAACACAGAACACTTAACGTGA 839
 QY 781 ACCAAGTCAAAGTAAATGACATTTAGTCTCTTGAAGACGCTTCAAACTGCTTTGAG 840
 Db 840 ACCAAGTCAAAGTAAATGACATTTAGTCTCTTGAAGACGCTTCAAACTGCTTTGAG 899
 QY 841 TCAAGGACATGTAAGATCTGATGAGCTTGTCTTTGATCCAGCAAAATCAAGTGAAC 900
 Db 900 TCAAGGACATGTAAGATCTGATGAGCTTGTCTTTGATCCAGCAAAATCAAGTGAAC 959

QY 901 AGCTAGAGGTGTGACAGTCCACACAGAGATCATTAACCACTTACCTTACTCAAT 960
 Db 960 AGCTAGAGGTGTGACAGTCCACACAGAGATCATTAACCACTTACCTTACTCAAT 1019
 QY 961 GTCTGAATTTGGAAGAACGAATGCTGCTATTAATTCCTCTTGGTATGCTTCAACCATTTG 1020
 Db 1020 GTCTGAATTTGGAAGAACGAATGCTGCTATTAATTCCTCTTGGTATGCTTCAACCATTTG 1079
 QY 1021 GGTACAGATTCAAGGCCAGAACCAAGTCCACACAGCAGCTCCGGAACTAGTCCAG 1080
 Db 1080 GGTACAGATTCAAGGCCAGAACCAAGTCCACACAGCAGCTCCGGAACTAGTCCAG 1139
 QY 1081 CCCGCAACTGCAACCAATCTTAAATAGACTCAATTTCTTGTGTTAGTCACTGCT 1140
 Db 1140 CCCGCAACTGCAACCAATCTTAAATAGACTCAATTTCTTGTGTTAGTCACTGCT 1199
 QY 1141 ACGAAAGTTGGGGAAGATATGTAATTCGAAGAAAAAGGCACTCTCGTTATGCTTTGC 1200
 Db 1200 ACGAAAGTTGGGGAAGATATGTAATTCGAAGAAAAAGGCACTCTCGTTATGCTTTGC 1259
 QY 1201 GAAAGTTTACCATCTGAAACTGTTAAAAATCTTGAAGCAAGTTATCAAAACAAGAGAG 1260
 Db 1260 GAAAGTTTACCATCTGAAACTGTTAAAAATCTTGAAGCAAGTTATCAAAACAAGAGAG 1319
 QY 1261 TGTTCACACACTTAACTGCTAAAAAGAAATGTGCTCTCGTGAACCAAGAAATTTTA 1320
 Db 1320 TGTTCACACACTTAACTGCTAAAAAGAAATGTGCTCTCGTGAACCAAGAAATTTTA 1379
 QY 1321 TGATTAAGCATTAATCTGTTAACTGAGGCTGATAAAGCTTGTGTAATTAAGGCTG 1380
 Db 1380 TGATTAAGCATTAATCTGTTAACTGAGGCTGATAAAGCTTGTGTAATTAAGGCTG 1439
 QY 1381 TAAATCTGATTTCCAGCCTTGAACAATTAAGAAGCTTGAATGATGATTCACATA 1440
 Db 1440 TAAATCTGATTTCCAGCCTTGAACAATTAAGAAGCTTGAATGATGATTCACATA 1499
 QY 1441 TAAAGAAAAATGTAAGATGATTTATTTGACATTCCTAGACCAATTAACCATCCAGAGG 1500
 Db 1500 TAAAGAAAAATGTAAGATGATTTATTTGACATTCCTAGACCAATTAACCATCCAGAGG 1559
 QY 1501 ACTTGGCAAAACCAAAATCTCAAAATGAGTATCTGAAGAGAAAGTTGCTCAATT 1560
 Db 1560 ACTTGGCAAAACCAAAATCTCAAAATGAGTATCTGAAGAGAAAGTTGCTCAATT 1619
 QY 1561 AGCTATTAAGTATCAACGTCAGATGCTTACATTTTGAATGAACATGATATACGTA 1620
 Db 1620 AGCTATTAAGTATCAACGTCAGATGCTTACATTTTGAATGAACATGATATACGTA 1679
 QY 1621 TGAAGGAGATGATATGTAACGCTCATATGAGGACATAGTCACTGATGGAAGAGATAG 1680
 Db 1680 TGAAGGAGATGATATGTAACGCTCATATGAGGACATAGTCACTGATGGAAGAGATAG 1739
 QY 1681 CCTTTCTGATTAAGAAAAAGTTGCAGCTCAAGCCTATATCAAAAGAAAAAGTATCCTTACC 1740
 Db 1740 CCTTTCTGATTAAGAAAAAGTTGCAGCTCAAGCCTATATCAAAAGAAAAAGTATCCTTACC 1799
 QY 1741 TCCATCTCCAGACGAGATGTTAAAGCAATCAACTGAGATAGTGCAGCAGCTATTTTA 1800
 Db 1800 TCCATCTCCAGACGAGATGTTAAAGCAATCAACTGAGATAGTGCAGCAGCTATTTTA 1859
 QY 1801 CAATCGTGTGAAGGGGAAAAAGCAATTCACCTGCTGCACTTCAATATATGTTGAGGA 1860
 Db 1860 CAATCGTGTGAAGGGGAAAAAGCAATTCACCTGCTGCACTTCAATATATGTTGAGGA 1919
 QY 1861 TACAGTTGAGTTAAAAAGGTAATTTGATTAATTCCTCATTAAGATCATTAACATATAT 1920
 Db 1920 TACAGTTGAGTTAAAAAGGTAATTTGATTAATTCCTCATTAAGATCATTAACATATAT 1979
 QY 1921 TAAATTTGCTGTTGATGATCAACATCAAAAGCTCCAAATGCTATACCTTGAAGA 1980
 Db 1980 TAAATTTGCTGTTGATGATCAACATCAAAAGCTCCAAATGCTATACCTTGAAGA 2039
 QY 1981 TTTGTTGAGAGATTAAGTACTAGTAACACCTGACGAAGCTCCACATTTAATGA 2040

```

Db      2040  TTTGTTTCGACGATTAAGTACTAGTAGAACACCCCTGACGAACTCCACATTTCTAATGA 2099
Qy      2041  TGGATGGGGCAATGCGACATGAGCATGTGTAGGCAAGAAAGCCACAGTGAAGATCCAAA 2100
Db      2100  TGGATGGGGCAATGCGACATGAGCATGTGTAGGCAAGAAAGCCACAGTGAAGATCCAAA 2159
Qy      2101  TAAGAACTTCAAAAGCGAGTGAAGAGCCAGTAGAGAGAAACACCTGTGAGCCAGAGTCCC 2160
Db      2160  TAAGAACTTCAAAAGCGAGTGAAGAGCCAGTAGAGAGAAACACCTGTGAGCCAGAGTCCC 2219
Qy      2161  TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAGAGAGAGAGAGTTTGTGCGAA 2220
Db      2220  TCAAGTAGAGACTGAAAAAGTAGAAGCCCAACTCAAGAGAGAGAGAGTTTGTGCGAA 2279
Qy      2221  AGTAACGAGATTTCTACTGTGAAGAGCCCAATGCAACAAAACTCTAGTGTGTTACGAAATTA 2280
Db      2280  AGTAACGAGATTTCTACTGTGAAGAGCCCAATGCAACAAAACTCTAGTGTGTTACGAAATTA 2339
Qy      2281  TTTGACTCTTCAAAATTTAGATACATAGTATCATGCGAGAGAGAGAGAGAGTTTGTGCG 2340
Db      2340  TTTGACTCTTCAAAATTTAGATACATAGTATCATGCGAGAGAGAGAGAGAGTTTGTGCG 2399
Qy      2341  GTTGTAAAGAGAGATTAATCTTCTCATCTGTAGTAGAGAGAGAGAGAGAGTTTGTGCG 2389
Db      2400  GTTGTAAAGAGAGATTAATCTTCTCATCTGTAGTAGAGAGAGAGAGAGAGTTTGTGCG 2448

```

RESULT 4
US-08-961-527-94
Sequence 94, Application US/08961527
Patent No. 6420135

GENERAL INFORMATION:

APPLICANT: Charles Kunesh
TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
NUMBER OF SEQUENCES: 391
CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville

STATE: Maryland
COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage

OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,527
FILING DATE:

CLASSIFICATION: 424
PRIOR APPLICATION DATA:

APPLICATION NUMBER:
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P1

TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 94:

SEQUENCE CHARACTERISTICS:
LENGTH: 8195 base pairs

TYPE: nucleic acid
STRANDEDNESS: double

TOPOLOGY: linear
US-08-961-527-94

Query Match 100.0%; Score 2388; DB 4; Length 8195;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

Qy      1  TTCTTACGAGTTGGAGCTGTATCAAGTAGAACCGTTAAGGAAATATATCGTTTCTTA 60
Db      3053  TTCTTACGAGTTGGAGCTGTATCAAGTAGAACCGTTAAGGAAATATATCGTTTCTTA 3112
Qy      61  TATATGATGAAAAACAAGCGACGCAAAAAACGGAGAAATTTGACTCTCTGATGAGTTAGCA 120
Db      3113  TATATGATGAAAAACAAGCGACGCAAAAAACGGAGAAATTTGACTCTCTGATGAGTTAGCA 3172
Qy      121  GCGGAAGGAATCAATGCTGAGCAATGCTGATCAAGATTAACAGCAAGGCTATGTTCAC 180
Db      3173  GCGGAAGGAATCAATGCTGAGCAATGCTGATCAAGATTAACAGCAAGGCTATGTTCAC 3232
Qy      181  TTCACATGGCGACCACTATCATTTATTAACAATGTAAGTTCTTATGACGCTATCATCAG 240
Db      3233  TTCACATGGCGACCACTATCATTTATTAACAATGTAAGTTCTTATGACGCTATCATCAG 3292
Qy      241  TGAAGAAATTAATCAATGAAGATCCAAACTATTAAGCTTAAAGATAGATATTTGTAATGA 300
Db      3293  TGAAGAAATTAATCAATGAAGATCCAAACTATTAAGCTTAAAGATAGATATTTGTAATGA 3352
Qy      301  GGTCAAGGATGATATGTTATCAAGATAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 360
Db      3353  GGTCAAGGATGATATGTTATCAAGATAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3412
Qy      361  TGCCCAAGCGGATTAAGCTCGCTACAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Db      3413  TGCCCAAGCGGATTAAGCTCGCTACAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3472
Qy      421  TCAACATGCTGAAGATGAGAACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 480
Db      3473  TCAACATGCTGAAGATGAGAACTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 3532
Qy      481  AGAGAGCTATATCAATGATATGTTATATCTTAAAGCTTCATATCATAGAGAGATAC 540
Db      3533  AGAGAGCTATATCAATGATATGTTATATCTTAAAGCTTCATATCATAGAGAGATAC 3592
Qy      541  TGGGATGCTTATATCTGTTCTCATGAGAGATCATTAACATTAATCTTAAGATGAGATT 600
Db      3593  TGGGATGCTTATATCTGTTCTCATGAGAGATCATTAACATTAATCTTAAGATGAGATT 3652
Qy      601  ATCAGCTAGCGAGTTGCTGCTGAGAGAGCTTCTATCTGCTGAGAGAAATCTGTCAA 660
Db      3653  ATCAGCTAGCGAGTTGCTGCTGAGAGAGCTTCTATCTGCTGAGAGAAATCTGTCAA 3712
Qy      661  TTCAAGAACCTATGCGGAGCAAAAAATAGCAATTAACCTTAAGAACAAATCGGATACCTC 720
Db      3713  TTCAAGAACCTATGCGGAGCAAAAAATAGCAATTAACCTTAAGAACAAATCGGATACCTC 3772
Qy      721  TGTAAAGATTCAGAGAACTACAAATACTAACACAGAGAACAAAGCAACTAACAGTCA 780
Db      3773  TGTAAAGATTCAGAGAACTACAAATACTAACACAGAGAACAAAGCAACTAACAGTCA 3832
Qy      781  AGCAAGTCAAGATATAGACATTTGATGCTTGAACAGCTTACAAACTGCTTTGAG 840
Db      3833  AGCAAGTCAAGATATAGACATTTGATGCTTGAACAGCTTACAAACTGCTTTGAG 3892
Qy      841  TCAACGACATGTAAATCTGATGCGCTTGTGATCCAGCAAAATCAAAATCGAAC 900
Db      3893  TCAACGACATGTAAATCTGATGCGCTTGTGATCCAGCAAAATCAAAATCGAAC 3952
Qy      901  AGTAGAGAGTTGTCAGTGCACACGAGAGATCATTAACACTTCACTTCACTCAAT 960
Db      3953  AGTAGAGAGTTGTCAGTGCACACGAGAGATCATTAACACTTCACTTCACTCAAT 4012
Qy      961  GTCTGAATTGGAAGAGAAATGCTGTATATTTCCCTTGTATGTTATCAAAACATTTG 1020
Db      4013  GTCTGAATTGGAAGAGAAATGCTGTATATTTCCCTTGTATGTTATCAAAACATTTG 4072
Qy      1021  GGTACCAATTCAGGCGCAAGAACCAAGTCCCAACCAAGCTCCGAACTTAATCCAGG 1080
Db      4073  GGTACCAATTCAGGCGCAAGAACCAAGTCCCAACCAAGCTCCGAACTTAATCCAGG 4132

```

```

QY 1081 CCCGCACTGCACCAATCTTAAATAGACTCAAACTCTTTGTTAGTCACTGCT 1140
DB 4133 CCCGCACTGCACCAATCTTAAATAGACTCAAACTCTTTGTTAGTCACTGCT 4192
QY 1141 AGGAAAGTTGGGGAAGATATGATTCGAGAGAAAGGGGATCTTCGTTATGCTTTG 1200
DB 4193 AGGAAAGTTGGGGAAGATATGATTCGAGAGAAAGGGGATCTTCGTTATGCTTTG 4252
QY 1201 GAAAGATTACCATCTGAAGCTGTAAATCTTGAAGAGATATCAAAACAAGAG 1260
DB 4253 GAAAGATTACCATCTGAAGCTGTAAATCTTGAAGAGATATCAAAACAAGAG 4312
QY 1261 TGTTCACACACTTTAACTGCTAAAAAAGAAATGTTGCTCTCTGACCAAGATTTTA 1320
DB 4313 TGTTCACACACTTTAACTGCTAAAAAAGAAATGTTGCTCTCTGACCAAGATTTTA 4372
QY 1321 TGATTAAGCATATATCTGTTAACTAGGCTCATTAAGCTTGTGMAATTAAGGCTG 1380
DB 4373 TGATTAAGCATATATCTGTTAACTAGGCTCATTAAGCTTGTGMAATTAAGGCTG 4432
QY 1381 TAATCTGATTTCCAGCCTTAGACAAATTAAGAAGCTTGAATGAAATGCACTAA 1440
DB 4433 TAATCTGATTTCCAGCCTTAGACAAATTAAGAAGCTTGAATGAAATGCACTAA 4492
QY 1441 TAAAGAAAAATTGATGATATTTATGTCATTCAGACCAATTAACCATCAAGCG 1500
DB 4493 TAAAGAAAAATTGATGATATTTATGTCATTCAGACCAATTAACCATCAAGCG 4552
QY 1501 ACTTGCAAAACCAATTTCTCAAATTGAGTACTGAAGCAAGATTGCTATTCAT 1560
DB 4553 ACTTGCAAAACCAATTTCTCAAATTGAGTACTGAAGCAAGATTGCTATTCAT 4612
QY 1561 AGCTGATAGTATACAGCTGATGCTTAATCTTTGATGAAACATGATATATCAGTA 1620
DB 4613 AGCTGATAGTATACAGCTGATGCTTAATCTTTGATGAAACATGATATATCAGTA 4672
QY 1621 TGAAGAGATGATATGATAGCTCTCATATGGCCATAGCTGATTTGAAAGATAG 1680
DB 4673 TGAAGAGATGATATGATAGCTCTCATATGGCCATAGCTGATTTGAAAGATAG 4732
QY 1681 CCTTTCTGATTAAGAAAAAGTTGACAGCTCAAGCTTACTTAAAGAAAAAGTATCTTACC 1740
DB 4733 CCTTTCTGATTAAGAAAAAGTTGACAGCTCAAGCTTACTTAAAGAAAAAGTATCTTACC 4792
QY 1741 TCCATCTCAGACGAGATGTTAAAGCAATCCAACTGAGATAGGACGACTATTTA 1800
DB 4793 TCCATCTCAGACGAGATGTTAAAGCAATCCAACTGAGATAGGACGACTATTTA 4852
QY 1801 CAATCGTGTGAAGGGGAAAAAGAAATTCACCTGTTGCACTTCCATATATGTTAGCA 1860
DB 4853 CAATCGTGTGAAGGGGAAAAAGAAATTCACCTGTTGCACTTCCATATATGTTAGCA 4912
QY 1861 TACAGTTGAGTTAAAAAGGTAATTTGATTTCTCATTAAGATCATTAACATATAT 1920
DB 4913 TACAGTTGAGTTAAAAAGGTAATTTGATTTCTCATTAAGATCATTAACATATAT 4972
QY 1921 TAAATTTGTTGTTGATGATCACAATCAAAAGTCCAAATGCTATATCTTGAAGA 1980
DB 4973 TAAATTTGTTGTTGATGATCACAATCAAAAGTCCAAATGCTATATCTTGAAGA 5032
QY 1981 TTTGTTTGGAGATTAAGTACTAAGTCAAGCACTGACGAAGTCCACATTTCTAATA 2040
DB 5033 TTTGTTTGGAGATTAAGTACTAAGTCAAGCACTGACGAAGTCCACATTTCTAATA 5092
QY 2041 TGAAGTGGGCAATGCGAGTGAAGTGTGTTAGCAAGAAAGACACAGTGAAGATCAAA 2100
DB 5093 TGAAGTGGGCAATGCGAGTGAAGTGTGTTAGCAAGAAAGACACAGTGAAGATCAAA 5152
QY 2101 TAAAGAACTTCAAAAGCGATGAAGACCAATTAAGGAAACACCTGCTGACCAAGATCCC 2160
DB 5153 TAAAGAACTTCAAAAGCGATGAAGACCAATTAAGGAAACACCTGCTGACCAAGATCCC 5212
QY 2161 TCAAGTAGAGACTGAAAAAGTGAAGCCCAACTCAAGAAAGCAGATTTTGTCTGCA 2220

```

```

DB 5213 TCAAGTAGAGACTGAAAAAGTGAAGCCCACTCAAGAAAGCAGAGTTTGTCTGCGAA 5272
QY 2221 AGTAACGATTTAGTCTGAAAGCCCAATGCAACAGAACTAGCTGTTTCAAAATA 2280
DB 5273 AGTAACGATTTAGTCTGAAAGCCCAATGCAACAGAACTAGCTGTTTCAAAATA 5332
QY 2281 TTTGACTCTTCAAAATTAATGATTAACATATGATATGATGAGAGAGCAAGAAATTAAGTTCG 2340
DB 5333 TTTGACTCTTCAAAATTAATGATTAACATATGATATGATGAGAGAGCAAGAAATTAAGTTCG 5392
QY 2341 GTTGTAAAGGAAGTAATCTTCACTGTAAGTAAGAAAAATTAAC 2389
DB 5393 GTTGTAAAGGAAGTAATCTTCACTGTAAGTAAGAAAAATTAAC 5441

RESULT 5
US-09-468-656A-11
; Sequence 11, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John E.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; CURRENT FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 2531
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-11

Query Match 41.5%; Score 991; DB 4; Length 2531;
Best Local Similarity 65.7%; Pred. No. 8.9e-263;
Matches 1609; Conservative 1; Mismatches 697; Indels 141; Gaps 6;

1 TTTCTTACGAGTTGGACCTGATCAAGCTTAAGGAAATTAATCGTTTC 57
DB 60 TTCTTATGATGCTTGAACGTTTACCAAGCTGTTGAGATTAAGAAAGTCTTAATCGAGTTGC 119
QY 58 CTATATAGATGAAACCAAGCGACGCAAAAAACGGAATTTGACTCTGATGAGGTTAG 117
DB 120 TTATATAGATGCTGATCAAGCTGCTGTCMAAAGCAGAAATTTGACACCAAGATGAGTCAAG 179
QY 118 CAAGGCTGAAGATGATGATGCTGAGCAAAATGCTCATCAAGATTAACAGACCAAGCTATGT 177
DB 180 TAAAGAGGAGGAGCAACGCGCAAAATTTGTTATCAAAATTAAGGATTAAGGATCAAGGTTATGT 239
QY 178 CACTTCAATGCGGACCACTATCAATTAATTAACAATGTAAGTTCCTTATGAGAGCTATCAT 237
DB 240 GACCTCTATGAGACCATTAATCAATTAATGATGAGCAAGGTTCTTATGAGGCAATCAT 299
QY 238 CAGTGAAGATTAATCAATGAAAGATCCAAACTATTAAGCTTAAAGATGAGATATTTGTA 297
DB 300 CAGTGAAGAGCTCTCATGAAAGATCCGAATTAATCAAGTTGAGAGATTCAGACATTTGTCAA 359
QY 298 TGAAGTCAAGGAGTGAATGTTATCAAGGTAGATGAAATTAATGTTTACTTAAAGA 357
DB 360 TGAATATCAAGGAGTGAATGTTATGATTAATGATTAACGTTAAATCTATGTTTAACTTAAAGA 419
QY 358 TGTCTGCCACGCGGATTAATGCTGCTTAACAAAGAGAAATCAATTCGACAAACCAAGAGCA 417
DB 420 TGCRCCTCATGCGGATTAATGATGAGCAAAAGAGATTAACGTTCAAGACAGAGAAAG 479
QY 418 TAGTCAACATCGTGAAGGTGAAGTCAAGAAACATGAGTGTGCTTGTGCTGCAAGTTC 477

```

Db 480 CAGTCATATCAT-----AACTCAAGCAGATATATGCTGTGTCAGCCAGAC 530
 Qy 478 GCAAGAGGCTTACTACAGATGATGTTATCTTTATGCTTCGATATCATAGAGA 537
 Db 531 CCAGAGGAGTTATACAGGATGATGGATATATCTTCAATGATCTGATATAGAGA 590
 Qy 538 TACTGTGATGCTTATATGCTTCTCATGAGATGATTAACATTAATTCTCTAAGATGA 597
 Db 591 CACGGGTATGCTTATATGCTTCTCATGAGGACCATTAACATTAATTCTCTAAGATGA 650
 Qy 598 GTTATACGTCAGGATGCTGCTGTCAGAGCCTTCTATGCTGTCAGAGAAATCTGTC 657
 Db 651 GTTATACGTCAGGATGCTGCTGTCAGAGCCTTCTATGCTGTCAGAGAAATCTGTC 695
 Qy 658 AAATCAAGAACCTATGCGCCGCAAAATAGCATTAACCTCAAGAACAACTGGGTACC 717
 Db 696 -GAGCAGAGATCTGCTCTTCTTCAAGTTCTATGATTAATGCAATCAAGTCAACCA 754
 Qy 718 TTCTGTAGCAATCCAGAACTACAAATTAACAACAAGCAACAACAGCACTAAGAG 777
 Db 755 GATGTGAGAGAACCAATCT-----GACTGTCACTCCAACTTA 794
 Qy 778 TCAGCAAGTCAAGTAATGACATGATATGCTTTGAAACGCTCTCAAACTGCTTT 837
 Db 795 TCATCAAAATCAAGGGGAAACATTTCAAGCCTTTAGCGTATGTAATGCTAAACCTT 854
 Qy 838 GAGTCAAGGACATGATGATGATGAGGCTGCTGCTTATGTCAGAGCAAAATCAGAGTCG 897
 Db 855 ATCAGAACGCAATGAGATCTGATGCGCTTATTTTCAACCGAGCAAAATCAGAGTCG 914
 Qy 898 AACAGCTGAGGCTGTGTCAGTCGCAACAGGATCATTAACACTTCACTCCCTACTCTCA 957
 Db 915 AACGCGCAGAGGTAGCTGCTCCCTCATGATACCATTAACCTTATCCCTTATGAGACA 974
 Qy 958 AATGCTGATTTGGAAGACGATGCTGCTGTTATTTCCCTGTTAGCTTAAACCA 1017
 Db 975 AATGCTGATTTGGAAGACGATGCTGCTGTTATTTCCCTGTTAGCTTAAACCA 1034
 Qy 1018 TTGGTATCAATTCAGGCGCAGAACCAAGTCCAAACCGATCCGGAACCTAGTCC 1077
 Db 1035 TTGGTATCAATTCAGGCGCAGAACCAAGTCCAAACCGATCCGGAACCTAGTCC 1094
 Qy 1078 AGGCGCGCAACCTGACCAAAATCTTAAATAGACTCAATTTCTTTGTTAGTACGCT 1137
 Db 1095 AAGTCCGCAACAGCTCCAGAACATCCAATTGATGAGGAA-----ATTGCTCAAGAAC 1148
 Qy 1138 GGTACGAAAGTTGGGGAGATATGATTCGAGAGAAAGGCGATCTCTGTTATGCTTT 1197
 Db 1149 TGTTCGAAAGTAGCGATGTTATGCTTTGAGAGAGATGAGATTTCTCGTTATATCCC 1208
 Qy 1198 TGCGAAGATTTACATCTGAACCTGTTAAAAATCTTGAAGCAAGTATACAAAACAGA 1257
 Db 1209 AGCCAGGATCTTTAGCAGAGAAACAGACAGGCAATGATAGCAAACTGGCCAGAGAGA 1268
 Qy 1258 GAGTGTTCACACACTTAACTGCTAATAAAAGAAATGTTGCTCTCTGTCAGCAAGATT 1317
 Db 1269 AAGTTTATCTATAGCTAGAGAACTAAGAAAACGACCTCCATCTATGATCGAGATT 1328
 Qy 1318 TTATGATTAAGCATTAATCTGTTAATCTGAGGCTCATAAAGCCTTTGTTGAAATAGGG 1377
 Db 1329 TTACATTAAGGCTTATGACTTACTAGCAAGAAATTCACCAAGATTTACTTGAATTAAGG 1388
 Qy 1378 TCGTAAATTCGATTTCCAGCCTTAGACAAATTAATAAGCTTGAATGATGATGCAC 1437
 Db 1389 TCGCAAGTTATTTGAGGCTTTGAGTAACCTGTTGAGACGACTCAAGATGCTCAAG 1448
 Qy 1438 TAATTAAGAAATTTGATAGATGATTTATTTGCAATTTCTACACCAATTAACCACTCAGA 1497
 Db 1449 TGATTAAGCTCAAGTATGAGAGATATCTTGCCTTTCTTACGTCGATTCGATCCAGA 1508
 Qy 1498 GCGACTTGGCAACCAATTTCTCAATTAAGTATATCTGAGAGCAAGATTTGCTATGCTCA 1557
 Db 1509 ACGTTTAGAAGAACCAATTAAGTACCACTGATGATGATGATTAAGTATGAGCA 1568

Qy 1558 ATTAGCTGATTAAGTATACAACTCAGATGTTACATTTTGTATGAATGATATATCAG 1617
 Db 1569 GTTGGCAGGCAAGTACACAGAGAGAGAGGTTATATCTTTATCTCTGATATATACAG 1628
 Qy 1618 TGATGAAGAGATCATATGTTAACGCTCATATATGAGGCTATGATGATTTGAGAGAA 1677
 Db 1629 TGATGAGGAGGATCTATATGTTAATCTCCACATATAGCCATATGATGATTTAAAGAA 1688
 Qy 1678 TAGCCTTTCTGATTAAGAGAAAGTTGACGCTCAAGCCTATATCTAAAGAAAGTATCCT 1737
 Db 1689 TAGTTTGTCTGAACCTGAGAGAGGCGGACCCAGCTTATGCTTAAGAGAAAGTTGAC 1748
 Qy 1738 ACCTCATCTCCAGACGAGATGTTAAAGCAATTCGAATCTGAGATATGTCAGACGAT 1797
 Db 1749 CCTCTCTTCGACACATCAGATCTGATTCAGAGAAATACAGGCAAAAGAGAGAGAT 1808
 Qy 1798 TTACATCTGTCGAAAGGGGAAACGAAATTCGACTGCTTGCATTCATATATGTTGA 1857
 Db 1809 CTACAACTGAGTGAAGAGCTTAAGAGGTCGACTGATGATGCTTCAATCTTCA 1868
 Qy 1858 GCATACGTTGAGTTAAACGCTAATTTGATATCTCTATATAGATCATTAACATTA 1917
 Db 1869 ATATACGTTGAGTCAAAACGCTAGTTTATATCATACCTTATATGACATTAACATTA 1928
 Qy 1918 TATTAATTTGCTGTTGATGATCAACATTAAGCTCAAAATGCTTATACCTTGA 1977
 Db 1929 CATCAATTTGAGTGTGTTGACGAAAGCCTTTATGAGGACCTTAAGGCTTATCTTGA 1988
 Qy 1978 AGATTTGTTGCGACGATTAAGTACTATGTAACACCTGTCAGAACGTCACATTTCAA 2037
 Db 1989 GGATCTTTTGGCAGCTGCAAGTACTATGTCGAACATTCGAACGATCGCATTTGAGA 2048
 Qy 2038 TGATGATGGGCAATGCGATGACATGCTTATGAGCAAGAAACCAACAGTGAAGATCC 2097
 Db 2049 TAAATGTTTGTATACCTGACGACATGTTCAAGAAACCAAAATGCTCAAGCTGATAC 2108
 Qy 2098 AATTAAGAACTTCAAGCGATGA----- 2121
 Db 2109 CAATCAACGGAAGAAACCAAGCGAGAGAAACCTCAGACAGAAACCTGAGAGAAAC 2168
 Qy 2122 -----AGAGCCAGT 2130
 Db 2169 CCTCGAAGAGAAACCGCAAGCAGAAACAGAGTCTCCAAACCAACAGAGAAAC 2228
 Qy 2131 AGAGAAACACCTGCTGAGCAGAGATCCCTCAAGTATGAGCTGAAAGATGAGAGCCA 2190
 Db 2229 AGAAGAAATCACAGAGAAATCAGAAAGAACTCAGGTCGAGCTGAAAGGTTGAAGAAA 2288
 Qy 2191 ACTCAAGAGCAGAGATTTGCTTGCAGAAAGTAAAGGATCTTCTGAAAGGCAATGC 2250
 Db 2289 ACTGAGAGAGGCTGAGATTTACTTGGAAAGATTCAGAGATTCATTAAGTCAATGC 2348
 Qy 2251 AACGAAGCTCTACCTGCTTTCAGAAATTAATTTGACTCTTCAATTAATGATTAACAATAG 2310
 Db 2249 CAAGAGAGCTCTCAGAGATTTAAATAATTTACTATTTGGAGCCAGGCAACAAATAC 2408
 Qy 2311 TATCATGCGCAGAGCAGAGAAATTAATTAATCTGCTGTTTAAAGAGATTA 2358
 Db 2409 TATATGCGCAGAGAGCTGAAAAAATTAATGCTTTTAAAGAGAGATTA 2456

RESULT 6
 US-09-468-656A-5
 ; Sequence 5, Application US/09468656A
 ; Patent No. 6582706
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John E.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
 ; FILE REFERENCE: 469201-444

CURRENT APPLICATION NUMBER: US-09/468,656A
 CURRENT FILING DATE: 1999-12-02
 PRIOR APPLICATION NUMBER: 60/113,048
 PRIOR FILING DATE: 1998-12-21
 NUMBER OF SEQ ID NOS: 14
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 5
 LENGTH: 2531
 TYPE: DNA
 ORGANISM: Streptococcus pneumoniae
 US-09-468-656A-5

Query Match 41.4%; Score 990; DB 4; Length 2531;
 Best Local Similarity 68.1%; Pred. No. 1,7e-262;
 Matches 1478; Conservative 0; Mismatches 636; Indels 57; Gaps 5;

QY 1 TTCTTACGAGTTGGAGCTGTATCAAGCTGAAAGGTTAAGAAA---TAACTGTTTC 57
 DB 60 TTCTTATGAACTTGTGCTGACCAAGCTGTGAGTTAAGAAAAGCTCTAATCGAGTTTC 119
 QY 58 CTATATAGATGAAAAACAAGCAGCAAGCAAAAACGGAATTTGACTCTGATGAGTTAG 117
 DB 120 TTAATATAGATGAGTATCAGGCTGTGTAAGAGCAAAAAGTTGACACCAAGATGAGTCA 179
 QY 118 CAAGCGTGAAGAAATCAATGCTGAGCAAAATGCTCATCAAGATTAACAGCAAGGCTATGT 177
 DB 180 TAAAGAGGAGGAGATCAACGCCAACAATCGTCATCAAGATTAACGATCAAGGTTATGT 239
 QY 178 CACTTCACTTGGCGACCACTATCATTTATCAATGTAAGGTTTCTTATGAGGCTATCAT 237
 DB 240 GACCTCTCAATGAGAACCAATATCATTAATATGCAAGGCTCCCTTATGATGCCATCAT 299
 QY 238 CAGTGAAGAAATTAATCATGAAAGATCCAAACTATAAGTAAAGATGAGATATTGTTAA 297
 DB 300 CAGTGAAGAGCTCTCATATAAAGATCCGAATTAATCAAGTGAAGATTCAGACATTTGCA 359
 QY 298 TGAGGTCAAGGGTGAATATGTTATCAAGGTAGATGAAAAATCTATGTTTACTTAAGGA 357
 DB 360 TGAATTCAAAGGGTGTATGTTATCAAGGTAGTGAATAATCTATGTTTACTTAAGGA 419
 QY 358 TCTGCGCCCGCGGATAGCTCCGTAACAAGAGAAATCAATCGACAAAACAAAGAGA 417
 DB 420 TGCAGCTCATGCGGATTAATTCGCAAAAAGAGATTTAAACGTCAGAGAGAGACA 479
 QY 418 TAGTCAACATGCTGAAGGTGAAGCTCCAGAAAAGATGAGTGTGCTTGTGCGACGTTTC 477
 DB 480 CAGTCAATATCAAGGGGGGTGTTCT---AACGATCAAGAGATGTTGACGCCAGAGC 533
 QY 478 GCAAGGAGCTTACTACAGATGATGTTATCTTTAATGCTTGTGATCATAGAGA 537
 DB 534 CCAAGGAGCTTATACAGAGATGATGTTATCTTCAATGATCTGATATCATTTGAGA 593
 QY 538 TACTGCTGATGCTTATATGTTCTCATGAGATCAATTAACATTAATCTCTAAGATGA 597
 DB 594 CAGGGGTGATGCTTATATGTTCTCATCGGCGACCAATTAATTAATCTTAAGATGA 653
 QY 598 GTTATCAGTACGAGTGTGCTGCTGCAAGAGCTTCCATCTGCTGAGAGAAATGTGTC 657
 DB 654 GTTATCAGTACGAGTGTGCTGCTGCAAGAGCTT----- 688
 QY 658 AAATTCAGAACTATCGCCGACAAATATAGCATTAACCTTCAGAGAAACCTGAGTACC 717
 DB 689 -----ATTGGAATGGAAGCAGGATCTGCTCTTCTTAAGTTCATGTTATTAATGC 740
 QY 718 TTCTGTAGCAATCCAGAACTACAAATTAATAACAGAACCAACAGAGACACTTAACG 777
 DB 741 AAATCAGCTCAACCAAGATGTCAGAGAACCAACATGATGATCACTCCAACTATCA 800
 QY 778 TCAAGAGTCAAGAAATGACATGATGCTCTTGAACAGCTCTAACAATGCTT 837
 DB 801 TCA---AAATCAAGGGGAAACATTTCAAGCTTTTACGTAATGTAATGCTTAACCTT 857
 QY 838 GAGTCAAGACATGATGAATCTGATGCTTGTGATTCAGACAAATCAACAGTGC 897

DB 858 ATCAGAGCCCATGTGGAATCTGATGCGCTTATTTTCAGCCAGGCAAAATCAAGTGC 917
 QY 898 AACACTGAGGTGTGTCAGTGCACACAGGAGATATTACCACTTACCTTACTCTCA 957
 DB 918 AACCCGAGAGGTGATGCTGCTCCATGATTAACATTACCACTTATCCCTTAAGACA 977
 QY 958 AATGCTGAAATGGAAGAAAGAAATGCTGATTAATCCCTTCTGTTAAGCAACA 1017
 DB 978 AATGCTGAATTTGAAAAAGAAATGCTGATTAATCCCTTCTGTTAAGCAACA 1037
 QY 1018 TTGGGTACAGATTCAGAGCCAGAACCAACCAAGTCCACACAGGACTCCGGAACCTAGTCC 1077
 DB 1038 TTGGGTACAGATTCAGAGCCAGAACCAACCAAGTCCACACAGGACTCCGGAACCTAGTCC 1097
 QY 1078 AGGCGCGACCTGACCAACCAATCTTAATTAATGACTCAAAATTTCTC-----TTT 1125
 DB 1098 AAGTCCGCAACCTGACCAAAATCTCAACAGCTCCAAATCAATGATGAGAAAT 1157
 QY 1126 GGTATGCTGAGTGTACGAAAGTTGGGAAAGATATGATTCGAAGAAAGGACATCTC 1185
 DB 1158 GGTCAAGAGAGTGTTCGAAAGTGAAGGATGATGATGCTTGAAGAGATGAGTTTC 1217
 QY 1186 TCGTATATCCAGCAAGAGATCTTACAGCAAAACAGCAGAGGATGATGAGCAACT 1277
 DB 1218 TCGTATATCCAGCAAGAGATCTTACAGCAAAACAGCAGAGGATGATGAGCAACT 1277
 QY 1246 ATCAAAACAAGAGATGTTTCAACACTTAACTGTAAGAAAGAAATGTTGCTCTCG 1305
 DB 1278 GGCACAGCAGAAATTTATCTCATAGCTAGAGCTTAAGAAATGACCTCCACTCTAG 1337
 QY 1306 TGACCAAGAAATTTATGATTAAGCATATATCTGTTAATGAGGCTCAAAAGCCTGTT 1365
 DB 1338 TGATGAGAAATTTATCAATAGCTTATGATCTTACAGCAAAATTCACCAAGATTTACT 1397
 QY 1366 TGNAAATAAGGGTCTAATCTGATTTCCAGCTTACAGCAAAATTTAGAACGCTTGA 1425
 DB 1398 TGATATTAAGGTTCACAAAGTGAATTTGAGCTTTGATTAACGTTGGAAGCACTCA 1457
 QY 1426 TGATGATGATTAATAAGAAATTTGATGATGATTTATTTGCACTTCTTACCACT 1485
 DB 1458 GATGATCCCAAGTGAATAAGTCAAGTGTGATGATATTTCTGCTTCTTGTGCTCGAT 1517
 QY 1486 TACCATCCAGAGGAGCTTGGCAACCAATCTCAATGATGATTAAGTATGAGAGAGAGT 1545
 DB 1518 TCGTATCCAGAGCTTTAGGAAACCAATGCGCAATTAATCACTACGATGAGAGAT 1577
 QY 1546 TCGTATGCTCAATGAGTGTATAGTATACAACTCAGATGTTTCAATTTTGTATGACA 1605
 DB 1578 TCAAGTACGAAAGTTGGCAGGCAATACACAAACAGAGCTTATATCTTTGATCTCG 1637
 QY 1606 TGATATTAATCAAGTGAAGAGATGATGATTAAGCTCATATGAGGCAATGATGACTG 1665
 DB 1638 TGATATTAACAGTGTGAGGGGAGTCTATGTAATCACTCAATATGACCATAGCACAGT 1697
 QY 1666 GATTGGAAGAATACCTTTCTGATTAAGGAAAGTTGAGAGTCAAGCTTAATCAATAA 1725
 DB 1698 GATTAAAAAGATGATGTTGTAAGCTGAGAGCGGAGCGAGGCTTAATGCTTAAGA 1757
 QY 1726 AAAAGTATCTTACCTCAATCTCCAGAGCAGATGTTAAAGCAAAATCCAACTGAGATG 1785
 DB 1758 GAAAGTTTGAACCTCTCTTGCACAGACATGAGATTCAGAAATATCTGAGGCAAAAG 1817
 QY 1786 TGCAGAGCTATTAACAATGCTGTGAAGAGGAAAGCAATTCACCTGTTGCACTTC 1845
 DB 1818 AGCAGAGCTATCTCAACCGGTAAGAGCTAAAGAGGCACTGATGATGCTGATGCT 1877
 QY 1846 ATATATGCTGAGCATACAGTGAAGTAAAGAGTAAATGATTAATCTCCATTAAGA 1905
 DB 1878 TTACATCTTCAATATCTGATGAAGTCAAAAGGATTTAATCATATCTCATTTATGA 1937
 QY 1906 TCATTAACATATATTAATTTGCTTGTGATGATCAACATCAAGTCCAAATG 1965


```

Db      1159 TCGTATATATCCAGCCAGAGATCTTTTACAGAGAAAGCAGCGAGCATGTATGCAAACT 1218
QY      1246 ATCAAAACAGAGAGTGTTCACACACTTAACTGCTAAAGAAATATGTCCTCTCG 1305
Db      1219 GGGCAAGCAGAGAAAGTTATCTATAGCTAGAGAGCTAAGAAACTGACCTCCACTG 1278
QY      1306 TGAACCAAGATTTTATGATTAAGCATATAATCTGTTAACTGAGGCTCATAAAGCTTGT 1365
Db      1279 TGATCGAGATTTTACATTAAGCTTATGACTAGCAAGAAATTCACCAAGATTTACT 1338
QY      1366 TGAATAAGAGGTCTGATTTCTGATTTCCAGCTTAGACAAATTTTAGAACCTTGAA 1425
Db      1339 TGATATAAAGGTCGACCAAGTATTTTGAGCTTTGGAATACCTGTGGAGACTCAA 1398
QY      1426 TGATGATCGACTAATAAAGAAATTTGGAGATGATTTATGCAATCTCTACACCAAT 1485
Db      1399 GGAATGTCNCAAGTAAAGTCAAGTTAGTGAGAAATTTCTTGCCCTTCTTACCTCGAT 1458
QY      1486 TACCCATCCAGAGCGACTTGGCAACCAATTTCTCAATTTGAGTATAGTAAAGCAAGT 1545
Db      1459 TCGTATCCAGAACGTTTGAAGAAACCAATGCGCAATTTACTACACTGATGATGAT 1518
QY      1546 TCGTATGCTCAATTAGCTGATTAAGTATACACGTCAGATGTTACATTTTGGATGA 1605
Db      1519 TCAGATGAGCCAAAGTTGGCAGGCAAGTACACACAGAAAGCGTTATATCTTGATCCTCG 1578
QY      1606 TGATATATCAGATGAAGAGATGATATATGTAAGGCTCATATGGCCATGTCATG 1665
Db      1579 TGATATATCAGATGAAGAGAGGATCTATGTAATCTCCATATATACCATAGCCATG 1638
QY      1666 GATTGGAAGATAGAGCTTCTGATTAAGGAAAGTTGAGAGCTCAAGCTTACTAAAG 1725
Db      1639 GATTAAAAAGATAGTTTGTCTGAGAGCTGAGAGAGGCGACCCAGGCTTATCTTAAG 1698
QY      1726 AAAAGGTATCTTACTCTTCCAGACGAGATGTTAAAGCAATTCACCTGAGATAG 1785
Db      1699 GAAAGGTTTGAAGCTTCTTCCAGACATCAGGATTCAGGAAATTCAGAGCAAAAG 1758
QY      1786 TCGAGAGCTATTTTACATCGTGTGAAGGGGAAAGCAATTCCTGCTGACTTCC 1845
Db      1759 AGCAGAGCTATCTTACACCGCGTGAAGAGCTTAAGAGTGCCACTGATGTATGCC 1818
QY      1846 ATATATGTTGAGCATACAGTTGAGGTTAAAGCGTAAATTTGATTTATCTCATAGA 1905
Db      1819 TTACATCTTCAATATACGTGAGAGTCAAAAGCGTATGTTATATATCTCATATAGA 1878
QY      1906 TCATTATCAATATTTAAATTTGCTGTTGATGATCAACATACAAAGCTTCAATAG 1965
Db      1879 CCATTACCAATTAACATAATTTGAGTGTGTCGAAAGGCTTTATGAGGCACTTAAG 1938
QY      1966 CTATACCTTGAAGATTTGTTGGCGAGATTAAGTACTAGTGAACACCTGAGAGAG 2025
Db      1939 GATATCTCTTGAAGATCTTTGGCGACTGTCAAGTCTATGTGCAACATCCAAACCAAG 1998
QY      2026 TCCACATTTTATGATGATGAGGCGAATGCGAGTGAAGTGTGTTGCGAAGAAAGCA 2085
Db      1999 TCCGCTTATGATATGTTTGTGTAACGTTAGCGCATGTTTCAAGAAACAAATAG 2058
QY      2086 CAGTGAAGATCCAAATTAAGAACTTCAAGCGATGAAGCCAGTGAAGAAACACTCG 2145
Db      2059 TCAGGTATACCAATCAACGAAACCAAGCGAGGAAACCTCAGACGAGAAAGCC 2118
QY      2146 TGAGCCAGAGTCCCTCAAGTAGAGA 2171
Db      2119 TGAGGAAGAAACCCCTCGAAGAGAGA 2144

```

RESULT 8
 US-09-536-784-65
 ; Sequence 65, Application US/09536784
 ; Patent No. 6573082
 ; GENERAL INFORMATION:

```

; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/536,784
; FILING DATE: 30-Oct-1997
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: OCT-30-1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Michelle S. Marks
; REGISTRATION NUMBER: 41,971
; REFERENCE/DOCKET NUMBER: PB340P3
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
;
; INFORMATION FOR SEQ ID NO: 65:
;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2290 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
; US-09-536-784-65
;
; Query Match      41.3%  Score 987.6; DB 4; Length 2290;
; Best Local Similarity 67.7%; Pred. No. 7.3e-262;
; Matches 1481; Conservative 0; Mismatches 645; Indels 60; Gaps 5;
;
;
; 1 TTCTTACGAGTGTGGAGCTGTATCAAGCTTAAGAGGTTAAGAAA---TAATCGTGTTC 57
; 4 TTCTTATGACTGTGGTCGTCACCAAGCTGTGATGAAGAAAGCTTAATCGAGTTTC 63
;
; 58 CTATATAGATGAAAAACAAGCAGCAAGCAAAAGCGAATTTGACTCTGATGAGGTAG 117
; 64 TTATATAGATGTGTATCAGGCTGTGTCAAAAGGCAAGAACTTGACACGATGAAGTCAG 123
;
; 118 CAAGGTGAAGAAATCAATGCTGAGCAAAATGCTCATCAAGATTAACAGACCAAGGCTATGT 177
; 124 TAAGAGGAGGAGATCAAGCGCAACAAATNTATCAAGATTACAGATCAAGGTTATGT 183
;
; 178 CACTTCACTGAGCAACCACTATCATTTATTAACAATGTAAGTTCTCTTATGAGCGTATCAT 237
; 184 GACCTCTATGAGACCATTTATCTATTAATGCAAGGTTCTTATGATGCCATCAT 243
;
; 238 CAGTGAAGATTTACTATGAAGATCAAACTTAAGCTTAAAGATGAAGATTTGTTAA 297
; 244 CAGTGAAGAGCTCTCATGAAGATCCGAATTAATCAGTTGAAGATTCAGACATTTGCA 303
;
; 298 TGAAGTCAAGGTTGATATGTTATCAAGGTAGATGAAGAAATCTATGTTTACTTTAAGA 357
; 304 TGAATAACAAGGTGTGTATGTATTAAGTAAACGTAATTAATCTATGTTTACTTTAAGA 363
;
; 358 TCGTGGCCAGCGGATTAAGCTCGTAAACAAAGGAAATCAATTCAGCAAAACAAAGAGA 417
; 364 TCGAGCTCATGCGGATTAATTTGAGCAAAAGAAAGATTAACCTCAGAAAGCAAGAGC 423
;
; 418 TAGTCAACATCGTGAAGGTGAAGCTCAAGAAAGATGAGTGTGTTGCTTGGCAGCTTC 477
; 424 CAGTCATATATAT-----ACCTCAAGAGCAGATTAATGCTGTTGCTGAGCGCAGAC 474

```

478 GCAAGACGCTAATACTACAGATGATGTTATCTTTATGCTTCTGATATCATAGAGA 537
 475 CCAAGAGCTTATACAGAGATGATGGATATCTTCAATGATCTGATATCATAGAGA 534
 538 TACTGATGATGTTATATGCTTCTCCTCAGAGATCATTAACCTTATCTTAAAGATGA 597
 535 CACGGGATGATCTTATATGCTTCTCAGGGCAGCATTAACCTTATCTTAAAGATGA 594
 598 GTTATACCTAGCGGCTTGGCTGCTGCAAGACCTTCTATCTGCTGAGAGAAATCTGTC 657
 595 GTTATACCTAGCGGCTTGGCTGCTGCAAGACCTTCTATCTGCTGAGAGAAATCTGTC 629
 658 AAATTCAGAACTCTGCGCCGCAAAATAGCGATTAACCTTCAAGAACAACTGGGTAC 717
 630 -----ATTGGAATGGGAAGCAGAGATCTGCTCTTCTTCAAGTCTTATGATTAATGC 681
 718 TTCTGTAAGCAATCCAGAACTCAAAATCTAACAAGCAACAAACAGCAACCTAACAG 777
 682 AAATCAGCTCAACCAAGATGTCAGAGAACCAATCTGACTGCTCACTCCAACTTATCA 741
 778 TCAGCAAGTCAAGATATGACATGATGATGCTTGAAGACGCTTCAAACTGCTT 837
 742 TCA---AAATCAAGGGGAAACATTTCAAGCTTTTACGTGATGATGCTAAACCTT 798
 838 GAGTCAAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 897
 799 ATCAGAACGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 858
 898 AACAGTAGAGGTGTCAGTCGACAGCAGAGATCATTAACCTTATCTTATCTTATCTCA 957
 859 AACCGCCAGAGGTGATGCTGCTCCTCATGTAACCATTAACCTTATCTTATGAGACA 918
 958 AATGCTCAATTTGGAAGAACAAATGCTGCTATTTATTTCCCTGCTTATCTTATGAGACA 1017
 919 AATGCTCAATTTGGAAGAACAAATGCTGCTATTTATTTCCCTGCTTATCTTATGAGACA 978
 1018 TTGGTATCAGATTCAGAGCCAGAACCAACAGTCCAGCAACCGACTCCGGAACCTAGTCC 1077
 979 TTGGTATCAGATTCAGAGCCAGAACCAACAGTCCAGCAACCGACTCCGGAACCTAGTCC 1038
 1078 AGGCCCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTCTC-----TTT 1125
 1039 AAGTCCGCAACCTGACCAAAATCTTCAACAGCTCCAGCAATCAATTTGATGAGAAAT 1098
 1126 GGTATAGTACGCTGATCAGAAAGTTGGGAGAGATATGATTTCAAGAAAGGGCATCTC 1185
 1099 GGTATAGTACGCTGATCAGAAAGTTGGGAGAGATATGATTTCAAGAAAGGGCATCTC 1158
 1186 TCGTTATGCTTTTGGGAAAGATTTTACATCTGAACCTGTTAAATCTTGAAGCAAGT 1245
 1159 TCGTTATGCTTTTGGGAAAGATTTTACATCTGAACCTGTTAAATCTTGAAGCAAGT 1218
 1246 ATCAAAACAGAGAGTGTTCACACATTTTACCTGCTTAAAGAAAGAAATGTTGCTCTCG 1305
 1219 GGCAGAGAGAGAAATTTATCTCATTAAGCTAGAGCTTAAGAAATGATGCTCCCATAG 1278
 1306 TGACCAAGATTTTATGATTAAGCATATATCTGTTAACTGAGGCTCATTAAGCTTGT 1365
 1279 TGATCGAAGATTTTATGATTAAGCATATATCTGTTAACTGAGGCTCATTAAGCTTGT 1338
 1366 TGNAAATTAAGGCTGATATCTGATTTCCAGCCCTTACAAATTAATTAAGAGCTTAA 1425
 1339 TGATTAATTAAGGCTGATATCTGATTTTGGGCTTGAATTAACCTGTTGAAGCAAGT 1398
 1426 TGATGAATGATTAATTAAGAAATTTGATGATGATTTATTTGATTTCTTGAAGCAAGT 1485
 1399 GGATGTCAGATGATTAAGTCAAGTTAGTGAGATATTTCTGCTTCTTGAAGCTGAT 1458
 1486 TACCATTCAGAGGACTTGGCAAACTTAATTTCTAAATTTGATTTACTGAGAGCAAGT 1545
 1459 TCGTATTCAGAGGCTTGAAGAAACCAAAATGCGCAATTAACCTGATGATGATGAT 1518

1546 TCGTATTCCTCAATTAAGCTGATTAAGTATACAACTGATGATGATTTTGAAGACA 1605
 1519 TCAGTATGCAAGGTTGGAGGCAAGTACAAAGAAAGGATATATCTTGAATCTTCC 1578
 1606 TGATATTAATCAGTATGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1665
 1579 TGATATTAACGATATGAGGGGATGCTTATGATTAACCTCAATATGACCTTACCTG 1638
 1666 GATTGGAAGATGAGCTTCTTCTGATTAAGAAAGGTTGAGCTCAAGCTTATCTTAAAGA 1725
 1639 GATTAAAGAAATGATGTTGCTGTAAGCTGAGAGCGGCGAGCTTATGCTTAAAGA 1698
 1726 AAAGGATTCCTTACCTCATCTCCAGACGAGATGTTAAAGCAATCCAACTGAGATAG 1785
 1699 GAAAGTTTGAACCTCTCTTCTGACAGACATCAAGATTCAGAAATTAAGGCAAAAG 1758
 1786 TGACAGAGCTATTTTCAATGCTGTTGAAGGGGAAAGGAAATTCATCTGTTGACATCC 1845
 1759 AGCAGAGCTATCTACACACGCGTGAAGAGAGCTTAAGAGGTCACCTGATGCTATGCC 1818
 1846 ATATATGTTGAGATACAGTGAAGTAAAGGTTAAAGGTTATTTGATTTCTCATTAAGA 1905
 1819 TTACAAATCTTCAATATATCTGATGAGATCAAAAGGTTATTAATCAATCTTATTAAGA 1878
 1906 TCATTACCAATATATTAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGAT 1965
 1879 CCATTACCAATATATCAATTTTGAAGTGTGTTGAAGAGGCTTTATGAGGACTTAAGGG 1938
 1966 CTATACCTTGAAGATTTGTTGCGAGATTAAGTACTAGTAAACACCTGAGCAAG 2025
 1939 GTATACCTTGAAGATCTTTTGGGAGCTGCAAGTACTAGTCAAACTCAAAAGCAAG 1998
 2026 TCCCATTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2085
 1999 TCCCATTTCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2058
 2086 CAGTGAAGATCCAAATTAAGACTTCAAAAGCGATGAAGAGCAGTAAAGCAACCTGTC 2145
 2059 TCAGTATGATTCATATCAATCAAAAGGAAACCAAGGAGAGAAACCTCAGACAGAAATCC 2118
 2146 TGAGCCGAAAGTCCCTCAAGTAAAGA 2171
 2119 TGAGGAAGAAACCCCTCAGAGAGAGA 2144

RESULT 9
 US-08-961-527-243
 ; Sequence 243, Application US/08961527
 ; Patent No. 6420135
 ; GENERAL INFORMATION:
 ; APPLICANT: Charles Kunsch
 ; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
 ; NUMBER OF SEQUENCES: 391
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/961,527
 ; FILING DATE:
 ; CLASSIFICATION: 424
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:

ATTORNEY/AGENT INFORMATION:
 NAME: Brookes, A. Anders
 REGISTRATION NUMBER: 36,373
 REFERENCE/DOCKET NUMBER: PB340P2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 181:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1342 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 US-08-961-083-181

Query Match 16.1%; Score 385.4; DB 3; Length 1342;
 Best Local Similarity 65.3%; Pred. No. 3.2e-96;
 Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

```

Oy 37 TAAAGAAATAATCGTGTTCCTATATAGATGAAACAGCGACGCAAAAACGGAGAA 96
Db 28 TAAAGCAATATATCTGTCTCTTATGTGATGGCAGCCAGTCAAGTCAGAAAAGTGA 87
Oy 97 TTTGACTCTGATGAGTTAGCAAGCGTGAAGAAATCAATGCTGACAAATCGTCATCA 156
Db 88 CTGACACCAAGCAGGTTAGCCAGAAAGAAAGAAATTCAGGCTGACAAATTTGATCA 147
Oy 157 GATAACAGACCAAGGCTATGTCACTTCACATGCGACCACTATCTTTTACAAATGGTAA 216
Db 148 AATTACAGATCAGGGCTATGTAACTCACCGTACCACCTATCTATTAATGGGAA 207
Oy 217 GGTTCCTATGACGCTATCATCAGTGAAGAAATTAATCAAGAAAGATCCAACTATAGCT 276
Db 208 AGTTCCTATGATGCTCTCTTTAGTGAAGAACTCTTGATGAAGATCCAACTATCACT 267
Oy 277 AAAAGATGAGATATTTGTAATGAGTCAAGGCTGATATGTTATCAAGATAGATGAAA 336
Db 268 TAAAGACCGTATATTTGCAATGAAGTCAAGGCTGTTATATCATCAAGGTGATGAAA 327
Oy 337 ATACATATGTTTACCTTAAGAGTCTGCCAGCGGATTAACCTCCGTACAAAAGAGAAAT 396
Db 328 ATATATATGTCTACCTGAAGATGACGCTCATCTATATGTTCCAACTAAAGATGAAT 387
Oy 397 CAATGACAAAACAAAGACATGCAATCATGTAAGGTGAATCCAGAAACGATGG 456
Db 388 CAATGCTAAAAACAAAGACATGTCAAAAGATATAGAAAGTTA-----ACTC 435
Oy 457 TCGTGTTCCTTGGCAGCTTCGCAAGACGCTATACTACAGATGATGTTATCTTTAA 516
Db 436 TAATGTCTGTAGCAAGGTCTCAGGAGATATACGAAATGATGTTATGCTTTAA 495
Oy 517 TCGTTCATATCATAGAGATCTGTGATGCTTATATCTTCTCTCATGAGATCATTA 576
Db 496 TCCACCTATATATTCGAAGATACGGGTAAATGCTTATATCTTCTCATGAGGCTACTA 555
Oy 577 CCATTACATTCCTTAAGATGAGTATCACTAGTGGAGTGGTGTGGAAGACCTTCT 636
Db 556 TCATCAATTCCTCAAAAGGATTTATCTGCTAGTAAATAGCAGAGCTTAAAGCATCT 615
Oy 637 ATCTGCGAGAAATCTGTCAAAATTCAGAACTATGCGCGACAAATAGCATATAC 696
Db 616 GGGTGAAGAAATATAGCAACGAGTCAGTAAAGTATCTTCAACAGCTAGTGAACA--- 672
Oy 697 TTCAAGAACAACTGGGTACTCTTGTAAAGCAATCCAGAACTACAAATACTACACAAG 756
Db 673 -----TAAACGCAATCTGT 687
Oy 757 CAACAACAGCAACTAAGTCAAGCAAGTCAAGTATGACATTTGATAGCTTTTAA 816
Db 688 AGCAAAAGATCAACTAGCAAGCCAGCAATTAATCTGAATCTCCAGAGCTTTTAA 747
Oy 817 ACAGCTTACAAACTGCTTTGATGCAAGCATAGATGATGAGCTTTGCTTTGA 876
  
```

Db 748 GAACTCTATGATTCACCTAGCCGCCAAGCTTACAGTGAATCAGATGAGCTGTCTTTGA 807
 Oy 877 TCCAGCAAAATCAACAGTGAACAGCTAGAGGTTGCGATGCGACAGGAGATCATTA 936
 Db 808 CCTGCTAAGATTAATCAGTGCATACCAAAATGAGTTGCCATTCGATGCGACCATTA 867
 Oy 937 CCATTCATCCCTTACTCTCAAAATGTGTAATGGAAGAAAGAAATGCTGATATATTC 996
 Db 868 CCACTTATCTCTTACAGCAAGCTTCTGCTTGAAGAAAGAAATGCCAAGATGTGCC 927
 Oy 997 CCTTCGT 1003
 Db 928 TATCAGT 934

RESULT 11
 US-09-536-784-181
 Sequence 181, Application US/09536784
 Patent No. 6573082
 GENERAL INFORMATION:
 APPLICANT: Choi et. al.
 TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
 NUMBER OF SEQUENCES: 452
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/536,784
 FILING DATE: 30-Oct-1997
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/961,083
 FILING DATE: OCT-30-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Michelle S. Marks
 REGISTRATION NUMBER: 41,971
 REFERENCE/DOCKET NUMBER: PB340P3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 181:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1342 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 SEQUENCE DESCRIPTION: SEQ ID NO: 181:
 US-09-536-784-181

Query Match 16.1%; Score 385.4; DB 4; Length 1342;
 Best Local Similarity 65.3%; Pred. No. 3.2e-96;
 Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

```

Oy 37 TAAAGAAATAATCGTGTTCCTATATAGATGAAACAGCGACGCAAAAACGGAGAA 96
Db 28 TAAAGCAATATATCTGTCTCTTATGTGATGGCAGCCAGTCAAGTCAGAAAAGTGA 87
Oy 97 TTTGACTCTGATGAGTTAGCAAGCGTGAAGAAATCAATGCTGACAAATCGTCATCA 156
Db 88 CTGACACCAAGCAGGTTAGCCAGAAAGAAAGAAATTCAGGCTGACAAATTTGATCA 147
Oy 157 GATAACAGACCAAGGCTATGTCACTTCACATGCGACCACTATCTTTTACAAATGGTAA 216
Db 148 AATTACAGATCAGGGCTATGTAACTCACCGTACCACCTATCTATTAATGGGAA 207
  
```

OY	217	GGTTCCTTAATGACGCATCATCGTGAAGAAATTACTCATGAAGAATCCAAACTATAGCT	276
Db	208	AGTTCCTTAATGATGCCCTCTTTAAGTAGAAGACTCTTGATGAAGAGATCCAACTATCAACT	267
OY	277	AAAAGATGAGATTAATTTGTAATGAGGTCAAGGGTGGATATGTTATCAAGGTAGATGAAA	336
Db	268	TAAAGACGGTATATTGTCATATGAAGTCAAGGGTGGTTATATCATCAAGGTGATGAAA	327
OY	337	ATACTATGTTTACCTTAAGAGATGCTGCCACGGGATTAACGTCCGTACAAAGAGAAAT	386
Db	328	ATATTATGCTACCTGAAAGAGTACAGTCAATGCTGTAATGTTTCAACTTAAATGATAAAT	387
OY	397	CAATGCAACAAAACAAGACATAGTCAACATCGTGAAGGGGAATCCCAAGAAACGATGG	456
Db	388	CAATGCTCAAAAACAAGACATGTCAAAGTAATAGGAAGTTA-----ACTC 435	
OY	457	TGCTGTTGCTTGGCAGCTTCCGAGAGACGCTATACTACAGATGATGTTATATCTTTAA	516
Db	436	TAAATGTTGCTTAGCAAGGTCTCAGAGGAGATTAACGACAAATGATGTTATGTTTAA	495
OY	517	TGCTTCTGATATCATAGAGATATCGGTGATGTTTATTCGTTCCCTCATGAGATCATTA	576
Db	496	TCACACTGATATTATCGAAGATACGGGTAAATGTTATATGTTTCCATGAGAGAGTACTA	555
OY	577	CCATTACATTCCTTAAGAAATGAGTTATCACGTACGAGTGGCTGCTGCAGAAACCTTCCT	636
Db	556	TCACTACATTTCCCAAAAGCGATTTATCTGCTAGTAATTAGCAGCAGCTTAAGACAATCT	615
OY	637	ATCTGCTCGAGGAATAATCTGTCAAATTCAGAAACCTATCGCCGACAAATAGCGATTAAC	696
Db	616	GGCTGGAAAAAATATGCAACCGAGTCAGTTAAGCTATTTCTTCAACAGTAGTACAA--- 672	
OY	697	TTCAAGAACAACTGGGTACCTTCTGTAAAGCAATCAGGAACCTACAAATATCTAACCAAG	756
Db	673	-----TAAACCGCAATCTGT 687	
OY	757	CAACACACAGCAACACTTAACAGTCAAGCAAGTCAAAATATGACATGTATAGTCTTGGAA	816
Db	688	AGCAAAAGSATTGACTAGCAGACCGCAAAATTAATGTGAAATCTCCAGAGTCTTTTGA	747
OY	817	ACAGCTCTPACAAATGCTCTTTAGTACGACAGCATGTAGAAATCTGATGGCCTTGTCTTGA	876
Db	748	GGAACTCTATGATTCACCTAGCGCCCAAGTTTACATGATCATGATGGCCTGTCTTTGA	807
OY	877	TCCACACAAATCACAGTGCAGACGCTAGAGGTGTGAGTGCCACACGAGATCATTA	936
Db	808	CCCTCTAAGATTATTCAGTGTACACAAATAGAGTTGGATTCGCCCATGGCCACACATTA	867
OY	937	CCACTTCATCCCTTACTCTCAAAATGCTGAATTTGGAAGAACGAATGCTGTATTTATCC	996
Db	868	CCACTTATTCCTTACAGCAAGCTTTCTGCTTAGAAGAAAAGATTGCCAGATTTGGTGCC	927
OY	997	CCTTGCT 1003	
Db	928	TATCAGT 934	

RESULT 12
 ; US-09-468-656A-7
 ; Sequence 7, Application US/09468656A
 ; Patent No. 6582706
 ; GENERAL INFORMATION:
 ; APPLICANT: Johnson, Leslie S.
 ; APPLICANT: Adamou, John E.
 ; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 ; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
 ; FILE REFERENCE: 469201-444
 ; CURRENT APPLICATION NUMBER: US/09/468,656A
 ; CURRENT FILING DATE: 1999-12-02
 ; PRIOR APPLICATION NUMBER: 60/113,048
 ; PRIOR FILING DATE: 1998-12-21

[illegible]

1 CLASSIFICATION: 424
2 PRIOR APPLICATION DATA:
3 APPLICATION NUMBER:
4 FILING DATE:
5 ATTORNEY/AGENT INFORMATION:
6 NAME: Brooke, A. Anders
7 REGISTRATION NUMBER: 36,373
8 REFERENCE/DOCKET NUMBER: PB340P1
9 TELECOMMUNICATION INFORMATION:
10 TELEPHONE: (301) 309-8504
11 TELEFAX: (301) 309-8512
12 INFORMATION FOR SEQ ID NO: 355:
13 SEQUENCE CHARACTERISTICS:
14 LENGTH: 973 base pairs
15 TYPE: nucleic acid
16 STRANDEDNESS: double
17 TOPOLOGY: linear
18
19 US-08-961-527-355

Query Match	16.0%	Score	381.4	DB	4	Length	973
Best Local Similarity	77.5%	Pred. No.	3.4e-95				
Matches	492	Conservative	0	Mismatches	131	Indels	12
						Gaps	2

QY	1	TTCTTCACGAGTGGGACCTGATACAGCTGAAAGCGTTAAGAAAA--TAATCGTGTTC	57
Db	157	TTCTTATGACCTTGGACGTTACCAAGCTGTCAGAGTAAGAAAGCTAATCGAGTTGC	216
QY	58	CTATATGATGAGAAAAAAGACGACGCAAAAAAGAGAAATTGACTCTCGATGAGGTTAG	117
Db	217	TTATATATGATGCTGATCAGAGCTGGTGTCAAAAGCGCAAAAACTTGACACCATGATGAAGTCAAG	276
QY	118	CAAGCGGTGAAGAAATCAATGCTGAGAGAAATGCTCATCAAGTTAACAGCAAGGCTATGT	177
Db	277	TAAAGAGGAGGGATCAAGCCGAAATTTGTATCAAGAAATTCGAGATCAAGGTTATGT	336
QY	178	CACCTGACATGGCGACACCATATCATTTATACATGTAAGTTAGTTCCTTATGACGCTATCAT	237
Db	337	GACCTGCAATGAGACATATATCATCTACTATATGCAAGGTTCTTATGATGCATCAT	396
QY	238	CAGTGAAGAATTACTATGAAAGATCCAAACTATATAGCTAAAGATGAGATATTGTTAA	297
Db	397	CAGTGAAGAGCTCCTCATGAAAGATCCGAATTTATGCTGAAGAGTTCAGCATTTGCCAA	456
QY	298	TGAGGTCAAGGGTGGATGTTATCAAGGTAGATGGAATAATGTTATGTTACTTAAGGA	357
Db	457	TGAAATCAAGGTTGGTTATGCTTATTAAGGTAAAGGTAAATPACTATGTTATTCCTTAAGGA	516
QY	358	TGCTGCCACGCGCGATAGATGCTCCGTACAAAAGAGAAATCAATGACAAAAACAGAGCA	417
Db	517	TGCAAGTCATGGGGATATATTTCCGACAAAABAAGATTAAAGCTCAGAGACAGAAAGC	576
QY	418	TAGTCAACATCGTGAAGGTGAACTTCAAGAAACGATGCTGCTGTTCCTTGGACAGTTC	477
Db	577	CAGCTCATATATCAT-----AAGTCAAGACCAATATATCTGTGTGTCGACGCAAGGC	627
QY	478	GCAGGACGCTTACTACTACAGATGATGGTTATATCTTTATGCTGTGATATACATAGAGGA	537
Db	628	CAAAGACGTTATACACGAGATGATGGGTATATCTTCATGATGATCTGATATCATTGAGGA	687
QY	538	TACTGTGATGCTTATATCGTTCTCATGAGATCATTTACATTACATTCCTTAAGATGA	597
Db	688	CAGGGGTGATGCTTATATATGCTTCTCAAGGACCATTTATACATTACATTCCTTAAGATGA	747
QY	598	GTTATCAAGCTAGGAGTTGGCTGTGCGAAGCCT	632
Db	748	GTTATCAGCTAGGAGTTAGCTGTGCGAAGCCT	782

RESULT 15
US-08-961-527-258
; Sequence 258, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:

APPLICANT: Charles Kunsch
 TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
 NUMBER OF SEQUENCES: 391
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/961,527

Query Match	10.2%	Score	243.4	DB	4	Length	1684	
Similarity	60.3%	Pred. No.	4.4e-57					
Best Local								
Matches 493;	Conservative	0;	Mismatches	226;	Indels	88;	Gaps	2;

QY	1629	ATCATATGTAAGCCCTCATATGCGGCATATGCTACCTGATTTGGAAAAATATGCCCTTTGTG	1688
Db	1	ATGCCATGTGTAACCTCCACATATGACCCCATAGCCACTGGAATTTAAAAAGATGTTGTCTG	60
QY	1689	ATAAGAAAAAGTTGCACTCAAAGCCTTATACTAAAGAAAAAGTATCCTACCTCATCTC	1748
Db	61	AACGTGAGAGAGGGGCA-CCGAGGCTTATGCTAAAGAGAAAGGTTTGACCCCTCCTTGA	119
QY	1749	CAGACGCAATGTTTAAAGCAATTCCAACTGGAGATAGTGCAGCGCTATTTTACATGCTG	1808
Db	120	CAAGCCATCGAGATTACGAAATATCTGAGCAAAAGAGAGAACTATCAACACCGC	179
QY	1809	TGAAAGGGGAAAAACGAAATTCCACTCGTGGACTTCCATATATGTTGAGCATACAGTTG	1868
Db	180	TGAAAGAGCTATGAAGAGTCCCACTGATGTGATGCTTACAACTTCAATATACTGTAG	239
QY	1869	AGGTTAAAAACGGTAAATTGATTAATTCCTCATAGAGATCATTTACATATATTAATTTG	1928
Db	240	AAGTCAAAAAACGGTAGTTTAACTATCAATCCTATTATGACCAATTCACATCAATCAAAATTTG	299
QY	1929	CTTGGTTTGATGATTCACATCAATCAAAAGTCCAAATGGCTATPACCTTGGAGATTTGTTTG	1988
Db	300	AGTGTGTTTGCAAGGCTCTTTATATAGGGCACTTAAGGGGTATATCTTGAAGATCTTTTGG	359
QY	1989	CGACGATTTAAGTACTACGTAGAAACACCCCTGACGAAGCTGCACATTTCAATGATGATGGC	2048
Db	360	CGACTGTCAAGTACTATGTGGAACATCCAAACGAACCTCGCAATTCAGATATAGGTTTGG	419
QY	2049	GCAATGCAAGTGAAGCATATGTTTAGCGAAGAAAGCCACAGTAAGATTCCAATTAAGAACT	2108
Db	420	GTAACGCTAGCGACCAATGTTCAAGAAACAAAAATGCTCAAGCTGATACCAATCAAAACGG	479

```

QY      2109 TCNAAGCGGATG-----2121
Db      480 AAAAACCAAGCAGAGAAACTCAGACAGAAAACTGAGGAGAAACCCCTGAGAG 539
QY      2122 -----AGGCCAGTAGGAGAAAC 2141
Db      540 AGAAACCCGAAAGCAGAAACGAGCTCCAAAACCAACGAGACACGAGAAATAC 599
QY      2142 CTGCTGAGCCGAAATCCCTCAAGTAGAGACTGAAAAAGTAAACCCCACTCAAGAG 2201
Db      600 CAGAGCAATTCAGAAACCTCAGCTGAGACTGAAAAAGTTGAAGAAAACTGAGAGAG 659
QY      2202 CAGAAATTTGCTGCGAAGTAGAGATTTAGTCTGAAAGCCCAATGCAACAGAACTC 2261
Db      660 CTGAAGATTTACTTGAAAAATCCAGATCCAAATTATCAAGTCCAAATGCAAGAGACTC 719
QY      2262 TAGCTGTTTACGAATAATTTGACTCTTCAATTATGATATCAATATGATCATGCGAG 2321
Db      720 TCACAGGATTTAAAAATATTTACTATTGTCAGCCAGACAAACAAATCTATTATGCGAG 779
QY      2322 AAGCAAAAAATTAATCTGCTGTTGTTAAAGAGATA 2358
Db      780 AAGTGAATAACTATTGCTTTATTAAAGAGAGATA 816

RESULT 16
US-08-743-637B-34/C
; Sequence 34, Application US/08743637B
; Patent No. 5994066
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: PICARD, Francois J.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIES-SPECIFIC AND UNIVERSAL DNA
; TITLE OF INVENTION: PROBES AND AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND
; TITLE OF INVENTION: IDENTIFY COMMON BACTERIAL PATHOGENS AND ASSOCIATED
; TITLE OF INVENTION: ANTIBIOTIC RESISTANCE GENES FROM CLINICAL SPECIMENS ...
; NUMBER OF SEQUENCES: 273
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 EAST WISCONSIN AVENUE
; CITY: MILWAUKEE
; STATE: WISCONSIN
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/743.637B
; FILING DATE: 04-NOV-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/526,840
; FILING DATE: 11-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586,90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 841 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:

```

```

; ORGANISM: Streptococcus pneumoniae
US-08-743-637B-34
Query Match      4.6%; Score 109; DB 2; Length 841;
Best Local Similarity 65.4%; Pred. No. 3e-20;
Matches 176; Conservative 0; Mismatches 90; Indels 3; Gaps 1;

QY      2090 GAGATCCAAATTAAGACTTCAAGCCGATGAGAGCCAGTAGAGAAAACTGCTGAG 2149
Db      633 GAGTCTCCAAACCAACGAGAGAACAGAGAAATATCCAGAGAAATCCAGAGAGA 574
QY      2150 CCAGAAGTCCCTCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGAGATT 2209
Db      573 TCAGAAAGACTCAGCTGAGACTGAAAGGTTAAGAA---AAGTAGAGAGAGCTGAAGAT 517
QY      2210 TTGCTTGGAAGTAGAGAGATTTAGTCTGAAAGCCCAATGCAAGAACTCTAGCTGCT 2269
Db      516 TTACTTGGAATAAATCCAGAAATTCATATCAAGTCCAAAGAGAGACTCTCAAGAGA 457
QY      2270 TTACGAATAATTTGACTCTTCAATATGATATCAATATGATATGATGAGAGAGAGAA 2329
Db      456 TTAATAAATTAATTTACTATTGTCAGCCAGAGCAACAATATATATGCGAGAGCTGAA 397
QY      2330 AAATTAATCTGCTGTTTAAAGAGATA 2358
Db      396 AACTATTGCTTTATTAAAGAGAGATA 368

RESULT 17
US-08-526-840B-34/C
; Sequence 34, Application US/08526840B
; Patent No. 6001564
; GENERAL INFORMATION:
; APPLICANT: BERGERON, Michel G.
; APPLICANT: OUELLETTE, Marc
; APPLICANT: ROY, Paul H.
; TITLE OF INVENTION: SPECIFIC AND UNIVERSAL PROBES AND
; TITLE OF INVENTION: AMPLIFICATION PRIMERS TO RAPIDLY DETECT AND IDENTIFY
; TITLE OF INVENTION: COMMON BACTERIAL PATHOGENS AND ANTIBIOTIC RESISTANCE GENES
; TITLE OF INVENTION: FROM CLINICAL SPECIMENS FOR ROUTINE DIAGNOSIS IN ...
; NUMBER OF SEQUENCES: 177
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: QUARLES & BRADY
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: Wisconsin
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/526,840B
; FILING DATE: 11-SEP-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/304,732
; FILING DATE: 12-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BAKER, Jean C.
; REGISTRATION NUMBER: 35,433
; REFERENCE/DOCKET NUMBER: 850586,90012
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5000
; TELEFAX: (414) 277-5591
; INFORMATION FOR SEQ ID NO: 34:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 841 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
US-08-526-840B-34

```

Query Match	Score	DB 3;	Length
4.6%;	109;		841

QY	2090	GAAATCCAAATAGAACTTCCAAAGCGATGAAAGCAGTATGAGAAACACTCTGAG	214
Db	633	GAGTCTCCAAACCAACAGAGGAAACAGAAAGAAATCACCAAGAAATCACCAAGGAA	574
QY	2150	CCAGAAGTCCTCAAGTNGAGACTAAAAAGTAAAGCCCACTAAAGAGCAGAATT	220
Db	573	TCAGAAGAACTCAGGTGAGACTGMAAGTTAAAG--AACTAGAGAGGCTTAAAGT	517
QY	2210	TTGCTTGGAAGTAACGGAATTCATGTGCGAAAGCCAAATGCACAAGAACTTACTGGT	226
Db	516	TTACTTGGAAAAATCCAGAAATCCAAATTCATCAAGTCCAAATGCCAAAGAGCTTTCACAGA	457
QY	2270	TTACGAATAATTTGACTCTTCAATTTATGATAACAAATAGTATCATGCGACAGAACAGA	232
Db	456	TTAAAAATAATTTACTATTTTGGCACCCAGGACACAAATATCATATATGCGAAGAACTGAA	397
QY	2330	AAATTAATCTGGCTTTAAAGAACTAA	2358
Db	396	AAATCATTTGGCTTTATTTAAAGAGGTAA	368

RESULT 18
US-08-232-463-14/c
; Sequence 14, Application US/08232463
Datafile: 070007

GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: PALCNER, P. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.2
CURRENT APPLICATION DATA:

```

; INFORMATION FOR SEQ ID NO: 14
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 7218 base pairs
;     TYPE: nucleic acid

```

```

; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-Fls
US-08-232-463-14

```

Query Match	2.4%;	Score 57.4;	DB 1;	Length 7218;
Best Local Similarity	3.4%;	Pred. No. 1.5e-05;		
Matches 13;	Conservative 224;	Mismatches 150;	Indels 0;	Gaps 0

[illegible]

```

RESULT 19
US-09-134-001C-2243
: Sequence 2243, Application US/09134001C
: Patent No. 6380370
: GENERAL INFORMATION:
: APPLICANT: Lym Doucelte-Stamm et al
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
: TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: GFC-007
: CURRENT APPLICATION NUMBER: US/09/134,001C
: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,964
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,779
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ. ID NOS: 5674
: SEQ. ID NO 2243
: LENGTH: 11091
: TYPE: DNA
: ORGANISM: Staphylococcus epidermidis
US-09-134-001C-2243

```

Query Match	2.2%;	Score 51.6;	DB 4;	Length 11091;
Best Local Similarity	47.3%;	Pred. NO. 0.00072;		
Matches 187;	Conservative	0;	Mismatches 205;	Indels 3;
			Gaps	1.

QY	Db	QY	Db
1096	8481	1156	8541
AAATCTAAATACACCAAAATCTCTTTGGTAGACGCTGGACCAAAAGTTGGGGA	AAAGCCAAAATGATGTAAATCAATCAATCAACGATTAACGAAGTTGAAATGCGAGCA	AGATATGATGTCCAGAAAAGGCGATCTCTCGTTATGTCTTTGGGAAA---GATTATCC	AAATAGTTTGATCAAAATCAATCACTTCAGACGAGATTTTGTATAAAAACGTATAGCAGT

EARLIER APPLICATION NUMBER: FR 95/07007
 EARLIER FILING DATE: 1995-06-13
 NUMBER OF SEQ ID NOS: 29
 SOFTWARE: Patent In Ver. 2.0
 SEQ ID NO 1
 LENGTH: 6152
 TYPE: DNA
 ORGANISM: P. falciparum
 US-08-973-462-1

Query Match 1.7%; Score 41.4; DB 3; Length 6152;
 Best Local Similarity 44.8%; Pred. No. 0.34;
 Matches 205; Conservative 0; Mismatches 247; Indels 6; Gaps 1;

1249 AAAACAAGAGAGTGTTCACACACTTAACTCTAAAAAGAAATGTGCTCTCGTGA 1308
 3097 AATATTGAAATTTAGAAAGAAATGCAGTAGAAAGTAATGAAATGTTGCGAGAAATTT 3156
 1309 CCAGAAATTTTATGATTAAGCATATATCTGTTAACTGAGGCTCATAAAGCCTTGTG 1368
 3157 AGAGAAATTTAAAGAACTGTATTTAATCTGATTAAGTAAGTAAGAAAGAGTGA 3216
 1369 AATATAGGCTCTAATTTCTGATTTCCAAAGCCTTAGACAAATTTAGAACGCTGAATGA 1428
 3217 AATTAGCGAGAAAGTTAGAAACAAATGAATGATTAAGCATTTTATAGTA 3270
 1429 TGAATCGACTAATAAGAAATTTGATGATTTATTTGGCATTTCTAGACCAATTAG 1488
 3271 AATATTGATATATGTAAGAAATCAAGAAATTTATTAACAGATATGTTCCAAATAT 3330
 1489 CCATCCAGAGGACTTGGCAACCAATTTCTCAATTTAGTATGTAAGACGAAGTTG 1548
 3331 AGAAACGAGTATGATTAATTCATCAGAAAGAAAGGTTGATTTGAATGAATGTTGTTAG 3390
 1549 TATTCCTCAATTTAGTATGATTAATCAACGTCAGATGTTATTTTATGTAACATGA 1608
 3391 TTGATTTTATGATTAATTAAGAAATTTAGAAAGAGTTATTTAATTAATTAAGAAATAT 3450
 1609 TATATCATGATGTAAGAGATGATATGTAACGCTCATATGAGCCATAGCATGAT 1668
 3451 TTCAAGTACTGAGAGTGTTCAGAAAGCTGTACTAAGCTAGAGCAAAATGTTATAT 3510
 1669 TGAAGAAATAGACCTTTCTGATTAAGAAAGTTGAG 1706
 3511 GGATGTTGATGTTCTGCTATGAAAGATCAATTTTATG 3548

RESULT 25
 US-09-461-697-193
 Sequence 193, Application US/09461697
 Patent No. 6277974

GENERAL INFORMATION:
 APPLICANT: COGENT NEUROSCIENCE, Inc.
 APPLICANT: Lo, Donald C.
 APPLICANT: Barney, Shawn
 APPLICANT: Thomas, Mary Beth
 APPLICANT: Portbury, Stuart D.
 APPLICANT: Putnam, Kasturi
 APPLICANT: Katz, Lawrence C.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
 TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
 TITLE OF INVENTION: CELL DEATH
 FILE REFERENCE: 10001-005-999
 CURRENT APPLICATION NUMBER: US/09/461,697
 CURRENT FILING DATE: 1999-12-14
 NUMBER OF SEQ ID NOS: 466
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 193

LENGTH: 696
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-09-461-697-193

Query Match 1.7%; Score 41.2; DB 3; Length 696;
 Best Local Similarity 53.0%; Pred. No. 0.13;
 Matches 88; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

2042 GGATGGGCAATGCCAGTACATGTGTAGGCAAGAAAGCAACAGTGAAGATCCAAAT 2101
 463 GGAAGTTGAAAGAGATGAAAGATGAAAGAGAGAGAGATGAAAGAGAGAGATGAAAGAT 522
 2102 AAGACTTCAAGAGGATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2161
 523 GTAAAGTCAAGAGAGATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 582
 2162 CAACTAGAGACTGAAAG 2207
 583 GAG 628

RESULT 26
 US-09-461-697-191
 Sequence 191, Application US/09461697
 Patent No. 6277974

GENERAL INFORMATION:
 APPLICANT: COGENT NEUROSCIENCE, Inc.
 APPLICANT: Lo, Donald C.
 APPLICANT: Barney, Shawn
 APPLICANT: Thomas, Mary Beth
 APPLICANT: Portbury, Stuart D.
 APPLICANT: Putnam, Kasturi
 APPLICANT: Katz, Lawrence C.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
 TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
 TITLE OF INVENTION: CELL DEATH
 FILE REFERENCE: 10001-005-999
 CURRENT APPLICATION NUMBER: US/09/461,697
 CURRENT FILING DATE: 1999-12-14
 NUMBER OF SEQ ID NOS: 466
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 191

LENGTH: 699
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-09-461-697-191

Query Match 1.7%; Score 41.2; DB 3; Length 699;
 Best Local Similarity 53.0%; Pred. No. 0.13;
 Matches 88; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

2042 GGATGGGCAATGCCAGTACATGTGTAGGCAAGAAAGCAACAGTGAAGATCCAAAT 2101
 466 GGAAGTTGAAAGAGATGAAAGATGAAAGAGAGAGAGAGATGAAAGAGAGAGAGATGAAAGAT 525
 2102 AAGACTTCAAGAGGATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2161
 526 GTAAAGTCAAGAGAGATGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAT 585
 2162 CAACTAGAGACTGAAAG 2207
 586 GAG 631

RESULT 27
 US-09-461-697-189
 Sequence 189, Application US/09461697
 Patent No. 6277974

GENERAL INFORMATION:
 APPLICANT: COGENT NEUROSCIENCE, Inc.
 APPLICANT: Lo, Donald C.
 APPLICANT: Barney, Shawn
 APPLICANT: Thomas, Mary Beth
 APPLICANT: Portbury, Stuart D.
 APPLICANT: Putnam, Kasturi
 APPLICANT: Katz, Lawrence C.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING

OY 2102 AAGAACTTCAAGCGATGAAGAGCCAGTAGAGGAAACCTCTGAGCCAGAACTCCCT 2161
 DB 130 GTAAAGTCTAAGAAAGATGAAAAAGAGAGAGAGATGAAAAAGAGATGAAAGTGAAT 189
 OY 2162 CAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAAGCAGAG 2207
 DB 190 GAGGAAGAACTGAAAAAGAGAAAGATTTAAAGAAAGAGAG 235

RESULT 31

US-09-461-697-184
 ; Sequence 184, Application US/09461697
 ; Patent No. 6277974
 ; GENERAL INFORMATION:
 ; APPLICANT: COGENT NEUROSCIENCE, Inc.
 ; APPLICANT: LO, Donald C.
 ; APPLICANT: Barney, Shawn
 ; APPLICANT: Thomas, Mary Beth
 ; APPLICANT: Portbury, Stuart D.
 ; APPLICANT: Puranam, Kasturi
 ; APPLICANT: Katz, Lawrence C.
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
 ; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
 ; FILE REFERENCE: 10001-005-999
 ; CURRENT APPLICATION NUMBER: US/09/461,697
 ; CURRENT FILING DATE: 1999-12-14
 ; NUMBER OF SEQ ID NOS: 466
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 184
 ; LENGTH: 1669
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; US-09-461-697-184

Query Match 1.7%; Score 41.2; DB 3; Length 1669;
 Best Local Similarity 53.0%; Pred. No. 0.2; 78; Indels 0; Gaps 0;
 Matches 88; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

OY 2042 GGATGGGCAATGCGAGTGCATGTGTAGGCAAGAAAGACCACAGTAGAATCCAAT 2101
 DB 665 GGAAGTGGAAAAAGAAATGAAATGAAAAAGAGAGAGATTAATAAGAGGGGAAAGAT 724
 OY 2102 AAGAACTTCAAGCGAGTAGAAGCCAGTAGAGGAAACCTCTGAGCCAGAACTCCCT 2161
 DB 725 GTAAAGTCTAAGAAAGATGAAAAAGAGAGAGAGATGAAAAAGAGATGAAAGTGAAT 784
 OY 2162 CAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAAGCAGAG 2207
 DB 785 GAGGAAGAACTGAAAAAGAGAAAGATTTAAAGAAAGAGAG 830

RESULT 32

US-09-627-122-21/c
 ; Sequence 21, Application US/09627122
 ; Patent No. 6472521
 ; GENERAL INFORMATION:
 ; APPLICANT: Uhlmann, Eugen
 ; APPLICANT: Greiner, Beate
 ; APPLICANT: Unger, Eberhard
 ; APPLICANT: Gothe, Gislinde
 ; APPLICANT: Schwerdel, Marc
 ; TITLE OF INVENTION: OLIGONUCLEOTIDES FOR THE INHIBITION OF HUMAN egs
 ; FILE REFERENCE: 02481.1678
 ; CURRENT APPLICATION NUMBER: US/09/627,122
 ; CURRENT FILING DATE: 2000-07-27
 ; NUMBER OF SEQ ID NOS: 23
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 21
 ; LENGTH: 5340
 ; TYPE: DNA

ORGANISM: Plasmodium falciparum
 US-09-627-122-21

Query Match 1.7%; Score 41; DB 4; Length 5340;
 Best Local Similarity 51.4%; Pred. No. 0.41;
 Matches 95; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

OY 191 GACCATCATCTATTATCAATAGTGAAGTTCCTATGACGCTATCATCAGTAGAATTA 250
 DB 740 GATCAAGTGCATCATTAATTAATTAATCTTATATATGATATCATTAAGACAAATTA 681
 OY 251 CTCATGAAGATCCAACTATAGCTTAAAGATGAGATATTTAATGAGTCAAGGT 310
 DB 680 TACTTGAAGATTAAATAATACCTGATGAAAGAGATAGAACAAATTAACAACATGAT 621
 OY 311 GGATATGTTATCAAGTAGATGAGAAATATCATGTTTAACTTAAGATGCGCCACCG 370
 DB 620 ATGATATTAAGTAAAGTATTTGAACCAATCGATGAAAACTTCCAAATTTCTATTAAGAT 561
 OY 371 GATTA 375
 DB 560 GATTA 556

RESULT 33

US-09-601-198-56/c
 ; Sequence 56, Application US/09601198
 ; Patent No. 6531583
 ; GENERAL INFORMATION:
 ; APPLICANT: Caselli, Gail H.
 ; APPLICANT: Chen, Elison Y.
 ; APPLICANT: Glass, Jennifer S.
 ; APPLICANT: Glass, John I.
 ; APPLICANT: Heiner, Cheryl R.
 ; APPLICANT: Leikowitz, Elliot
 ; TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
 ; FILE REFERENCE: UAB-13452/22
 ; CURRENT APPLICATION NUMBER: US/09/601,198
 ; CURRENT FILING DATE: 2000-12-08
 ; PRIOR APPLICATION NUMBER: 60/073,189
 ; PRIOR FILING DATE: 1998-01-30
 ; NUMBER OF SEQ ID NOS: 181
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 56
 ; LENGTH: 14066
 ; TYPE: DNA
 ; ORGANISM: Ureaplasma urealyticum
 ; US-09-601-198-56

Query Match 1.7%; Score 41; DB 4; Length 14066;
 Best Local Similarity 46.1%; Pred. No. 0.67; 160; Indels 0; Gaps 0;
 Matches 137; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

OY 2079 AAGACCAAGTGAATCCAAATTAAGAACTTCAAGCGAGTAGAAGCCAGTAGAGAAA 2138
 DB 2789 AAGATGATTAATGATGTTTAAATTAATTAATGATGACACTGTTAAATTAAGATGA 2730
 OY 2139 CACCTGCTGAGCCAGGAAGTCCCTCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAG 2198
 DB 2729 ATAAATTAATGAACAAAAATAGTTGTTAATTAAGAAATTAATGATGATTAATCAAG 2670
 OY 2199 AAGGAGAAATTTCTGCGGAAATGAATGCGATTCTAGTCGAAGCCAAATGCAACAGAA 2258
 DB 2669 GTCAATTCGATTAATTTAATTCAGAACTAAATTAATTAAGAAATTAATTAAGTA 2610
 OY 2259 CTAGCTGTTTGAATTAATTTGACTCTTCAATTAATGATTAATGATTAATGATG 2318
 DB 2609 AACCATTAATAACACACTTAATTTATCTGATGATTAATGATTAATGATTAATGAT 2550
 OY 2319 CAGAGAGAAAAATTAATCTGCTGTTGAAGAGAGATATCTTCATCTGTAAGTA 2375
 DB 2549 TAATTAAGTAACTGGAATCCGGTTTAAAGTTAATTAATCAACTCAAAATGATACTA 2493

RESULT 34

US-09-134-001C-925
 ; Sequence 925, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 925
 ; LENGTH: 1704
 ; TYPE: DNA
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-925

Query Match 1.7%; Score 40.6; DB 4; Length 1704;
 Best Local Similarity 44.6%; Pred. No. 0.29;
 Matches 160; Conservative 0; Mismatches 199; Indels 0; Gaps 0;

QY 1370 AATTAAGGTCGTAATTCGATTTCCAGCCTTAGACAAATTATAGACGCTTGAATGAT 1429
 DB 1114 ATTGAAGGTCGTAATAAATGATGACCAATCGCTGATATGTTTAAATATATTTTAA 1173
 QY 1430 GAATGACGTAATAAGAAAAATTTGTAGATGATTTATTTGGCATTTCTTACACCAATTAAC 1489
 DB 1174 GAATCTAATAGATTAAGATCATTTGGTTACAGACTTATTAAGATTATCAATATAGAACG 1233
 QY 1490 CATCAGACGCACTTGCCAAACCAATTTCTCAATAGATATAGACGCAAGTTGCT 1549
 DB 1234 CAAAAAGAACTTGAATAATATACATGATTTATCTGAATTTGCTTATTAATATAGAT 1293
 QY 1550 ATTGCTCAATTCGATATAGATACAGCTCAGATGTTTACATTTTGTATGAACATGAT 1609
 DB 1294 AATTGCAAAACCAAGCATATACATAGAGATCAATCAATCAATCGAAATTCGAAAAAGAT 1353
 QY 1610 AATTAAGGTAAGAGAGATGATATGTAAGCCCTCATATGGCCATAGCTCAGAT 1669
 DB 1354 GTCAATCATTTGAGGACATATAAATAATAGCGCAAGTTATTAATTTGCTATCAAT 1413
 QY 1670 GGAATAAGATAGCCTTTCTGATAAGAAAAAGTTGACGCTCAAGCCTATATCTAAGAAA 1728
 DB 1414 GGTATTAATATTTCTTCAGAGATATAATAGTAAATGTATATAGAAATGACAA 1472

RESULT 35

US-09-328-352-2472
 ; Sequence 2472, Application US/09328352
 ; Patent No. 6562958
 ; GENERAL INFORMATION:
 ; APPLICANT: Gary L. Breton et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
 ; FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC99-03PA
 ; CURRENT APPLICATION NUMBER: US/09/328,352
 ; CURRENT FILING DATE: 1999-06-04
 ; NUMBER OF SEQ ID NOS: 8252
 ; SEQ ID NO 2472
 ; LENGTH: 792
 ; TYPE: DNA
 ; ORGANISM: Acinetobacter baumannii
 US-09-328-352-2472

Query Match 1.7%; Score 40.2; DB 4; Length 792;
 Best Local Similarity 50.8%; Pred. No. 0.26;

Matches 121; Conservative 0; Mismatches 114; Indels 3; Gaps 1;

QY 1311 AAGAAATTTATGATTAAGCATATATCTGTTAAGGCTCATATAAGCCTGTTGMA 1370
 DB 95 AAGAACTTATTAAGCTTGACATTAATTTGATTCCTGATTTACCGGTGTTTGACAAC 154
 QY 1371 ATAAGGTCGTAATTCGATTTCCAGGCTTAGACAAATTATTAAGACGCTTGAATGAT 1430
 DB 155 TTAAGACGTCATAAATTTGTTGCCGATTTTA--AAGATACCATCGCTGATTTGCTG 211
 QY 1431 AATGCACTAATAAGAAAAATTTGTAGATGATTTATTTGGCATTTCTTACGACCAATTAAC 1490
 DB 212 AGTCGACTAATAATTCGCTTTATCTGATGATTAATTCGGAACATGACATTTTAACGC 271
 QY 1491 ATCCAGGCACTTGGCAACCAATTTCTCAATGATGATTAATCAAGCAAGTTG 1548
 DB 272 GTATGAACCACTTAGACAGCAAACTTTAAATGATGATTTGGCAGATGAATCTTG 329

RESULT 36

US-09-134-001C-322
 ; Sequence 322, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 322
 ; LENGTH: 30549
 ; TYPE: DNA
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-322

Query Match 1.7%; Score 40; DB 4; Length 30549;
 Best Local Similarity 48.1%; Pred. No. 1.9;
 Matches 112; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 1217 GAACTGTTAAATCTTGAAGCAAGTTATCAAAACAGAGAGGTTTCACACCTTTA 1276
 DB 10594 GCACCAATTTCAAAATTAATACCTTAACGTCAAAAGCAGACAGATTAACAAATGCA 10653
 QY 1277 ACTGCTAAAAAGAAATGTTGCTCTGTCGTCGACCAAGATTTTATGATTAAGCATATAT 1336
 DB 10654 AATTAATAATTTATTAAGAAAGCTTCAACCTTAGTACCAAGATGTCGATGAGAAATCTAAG 10713
 QY 1337 CTGTTAAGGAGGCTCATATAAGCCTTTGTTGAAATTAAGGCGTATATCTGATTTCCA 1396
 DB 10714 GTAGACGAAGCACTAGGATTAAGCAACGCCAATCAGCCTTAGAGCTATTAACAA 10773
 QY 1397 GCCTTAGCAATTAATTAAGAACGCTTGAATGATGATGCACTAATTAAGAAAA 1449
 DB 10774 GAATTAACGCAAGCATATTAATGATGATTAACCAACGAGATTTAAATATTA 10826

RESULT 37

US-09-134-001C-1027/C
 ; Sequence 1027, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; FILE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13

PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 1027
LENGTH: 297
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1028

Query Match 1.7%; Score 39.8; DB 4; Length 297;
Best Local Similarity 48.5%; Pred. No. 0.2;
Matches 110; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 1704 CAGCTCAAGCCTATCTAAAGAAAAAGGATCTCTACCTTCATCTCCAGACGCAATGTTA 1763
DB 281 CTGATCAAGCCGGTCTACGGTTAACTGATCTGCTCCAAATCAAGACGCTGAAGCAA 222
QY 1764 AAGCAATCCAACTGAGATAGTGACAGAGCTATTACATCTGTGAAGGGGAAAAAC 1823
DB 221 CCCCTAATCCGATATAAATATATACCACTAATGTGTCAAGACGGGATCAACAAATC 162
QY 1824 GAATTCACCTGCTTCGACTTCATATATGTTGACATACAGTTGAGGTTAAAAACGGTA 1883
DB 161 AAAGTACACAGATGATATATGATATCAAAATAGCCAAAGGTAACTAAACAAACA 102
QY 1884 ATTGATATTCTCTATAGGATCTTACCATATATTAAATTGCT 1930
DB 101 ATCAAAATGTGAACAGTAACTACTGGCGGTACAGATAAAGATGCT 55

RESULT 38

US-09-134-001C-1028
Sequence 1028, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 1028
LENGTH: 1716
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1028

Query Match 1.7%; Score 39.8; DB 4; Length 1716;
Best Local Similarity 48.5%; Pred. No. 0.49;
Matches 110; Conservative 0; Mismatches 117; Indels 0; Gaps 0;

QY 1704 CAGCTCAAGCCTATCTAAAGAAAAAGGATCTCTACCTTCATCTCCAGACGCAATGTTA 1763
DB 1322 CTGATCAAGCCGGTCTACGGTTAACTGATCTGCTCCAAATCAAGACGCTGAAGCAA 1381
QY 1764 AAGCAATCCAACTGAGATAGTGACAGAGCTATTACATCTGTGAAGGGGAAAAAC 1823
DB 1382 CCCCTAATCCGATATAAATATATACCACTAATGTGTCAAGACGGGATCAACAAATC 1441
QY 1824 GAATTCACCTGCTTCGACTTCATATATGTTGACATACAGTTGAGGTTAAAAACGGTA 1883
DB 1442 AAAGTACACAGATGATATATGATATCAAAATAGCCAAAGGTAACTAAACAAACA 1501
QY 1884 ATTGATATTCTCTATAGGATCTTACCATATATTAAATTGCT 1930
DB 1502 ATCAAAATGTGAACAGTAACTACTGGCGGTACAGATAAAGATGCT 1548

RESULT 39
US-09-601-198-3
Sequence 3, Application US/09601198
Patent No. 6531583
GENERAL INFORMATION:
APPLICANT: Cassell, Gail H.
APPLICANT: Chen, Elison Y.
APPLICANT: Glass, Jennifer S.
APPLICANT: Glass, John I.
APPLICANT: Heiner, Cheryl R.
APPLICANT: Letkowitz, Elliot
TITLE OF INVENTION: NUCLEIC ACID PROBES AND METHOD FOR DETECTING UREAPLASMA
TITLE OF INVENTION: UREALYTICUM
FILE REFERENCE: UAB-13452/22
CURRENT APPLICATION NUMBER: US/09/601,198
CURRENT FILING DATE: 2000-12-08
PRIOR APPLICATION NUMBER: 60/073,189
PRIOR FILING DATE: 1998-01-30
NUMBER OF SEQ ID NOS: 181
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 3
LENGTH: 1308
TYPE: DNA
ORGANISM: Ureaplasma urealyticum
US-09-601-198-3

Query Match 1.7%; Score 39.6; DB 4; Length 1308;
Best Local Similarity 46.8%; Pred. No. 0.48;
Matches 123; Conservative 0; Mismatches 140; Indels 0; Gaps 0;

QY 1202 AAAGATTACCATGTGAACGTGTTAAATCTTGAAGCAAGTATCAAAACAGAGAGT 1261
DB 184 AAAGATTACCATGTGTTAAATATAGTGTGATTTGCAAAATGGAAGTATACCTAAA 243
QY 1262 GTTTCACACCTTTACTGCTTAAAGAAAAAGTCTCTCTCTGACCAAGAAATTTAT 1321
DB 244 ACATTAGGAAGTAAAGATGATTTATACATATATGTTCTACCAACTAATTCAGAT 303
QY 1322 GATTAAGCATATATCTGTTACTGAGGCTCATTAAGCTTGTGTAATTAAGGTCGT 1381
DB 304 TATCAAAAATGGAATACAGTGCAGTAATATATTAATGAAGAAATTAATTA 363
QY 1382 AATCTGATTTCCAGGCTTATAGACAAATTTATAGAGCTTGAATGATCGACTAAT 1441
DB 364 AATCTGATTTTAAAGAGTGAAGAAATTTGTGACAGCTAATGATGAATGAATTT 423
QY 1442 AAAGAAAAATGTGATGATTT 1464
DB 424 GAATTTTAGGGGAAGTTAATTT 446

RESULT 40

US-08-916-421B-1/c
Sequence 1, Application US/08916421B
Patent No. 6503729
GENERAL INFORMATION:
APPLICANT: Bult et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococcus
TITLE OF INVENTION: jannaschii
FILE REFERENCE: PB275
CURRENT APPLICATION NUMBER: US/08/916,421B
CURRENT FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: Patentin version 3.1
SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Methanococcus jannaschii

FEATURE:
NAME/KEY: misc_feature
LOCATION: (28222) .. (28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (28257) .. (28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84773) .. (84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84808) .. (84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84812) .. (84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98120) .. (98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98159) .. (98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98239) .. (98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98266) .. (98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98343) .. (98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (103998) .. (103998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (148948) .. (148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (163385) .. (163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191989) .. (191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (231980) .. (231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234187) .. (234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234220) .. (234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234814) .. (234814)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309398) .. (309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309418) .. (309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312837) .. (312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312993) .. (312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (319226) .. (319226)
OTHER INFORMATION: n equals a, t, c, or g

NAME/KEY: misc_feature
LOCATION: (559167) .. (559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559241) .. (559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (600992) .. (600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (622708) .. (622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657081) .. (657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657203) .. (657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (674435) .. (674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (682442) .. (682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (713652) .. (713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (741684) .. (741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779455) .. (779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779676) .. (779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (855539) .. (855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (871619) .. (871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1084830) .. (1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1096846) .. (1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1119881) .. (1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1130881) .. (1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1310988) .. (1310988)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1313224) .. (1313224)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349473) .. (1349473)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1349491) .. (1349491)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1470091) .. (1470091)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1569020) .. (1569020)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature

```

LOCATION: (1602912)..(1602912)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1603734)..(1603734)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1637998)..(1637998)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc feature
LOCATION: (1664854)..(1664855)
OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

```

```

Query Match      1.6%; Score 39.4; DB 4; Length 1664976;
Best Local Similarity 45.5%; Pred. No. 21;
Matches 179; Conservative 0; Mismatches 211; Indels 3; Gaps 1;

```

```

Oy 1517 TCTCAATTGAGTATCTGAAAGCAAGTTCGATTGCTCAATTAGCTGATTAAGTATACA 1576
Db 1076571 TTTGGAAATTTTAAATAGCAAAATGCAATTTATATGCTAAATCATCATATTGATAAA 1076512

Oy 1577 ACGTCAGATGTTAATTTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 1636
Db 1076511 ATCTGAGATGAGAAAGATTAATTTGAACCTTTTATCTCTGTGAGATTATTAATGAGTTT 1076452

Oy 1637 GTAACGCTCATATGCGCATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1696
Db 1076451 CTATCTGATTTATTTGGCATTTGTTGAGAGAAATAGTTATTAATTCATTTGCAATGSCA 1076392

Oy 1697 AAAGTTGACGTCA--AGCTTATCTAAGAAAGGATTCCTACCTTCATCTTCAGAC 1753
Db 1076391 AAATATCCCGTAAAGAGAGCTGTGCGCAATTTTCAAGAGTTATCCATTAATTCATTAAT 1076332

Oy 1754 GCAGATGTTAAGCAATTCACATGAGATGATGATGATGATGATGATGATGATGATGATGAT 1813
Db 1076331 GGTGGCTCATATCATTTTAAACAGTAAATGAGGCTATATCTAATTAATGCGGAT 1076272

Oy 1814 GGGGAAAAAGCAATTCACATGCTGCTGATGATGATGATGATGATGATGATGATGATGATGAT 1873
Db 1076271 GTTTCATTTCCATATGCTTTATTTATGCTATTTCCATTAATCTACTCATCAAAATGSGG 1076212

Oy 1874 AAAAAGGTAATTGATTATTCCTCATAGAT 1906
Db 1076211 ATAAAAATAATCAAAAGATTTCCTCAAGCAT 1076179

```

```

RESULT 41
US-09-671-317-14/c
Sequence 14, Application US/09671317
Patent No. 6528260

```

```

GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marta
APPLICANT: Chumakov, Ilya
APPLICANT: Bouguetelret, Lydie
TITLE OF INVENTION: BIALLELIC MARKERS RELATED TO GENES INVOLVED IN DRUG METABOLISM
FILE REFERENCE: 62 US3 CIP
CURRENT APPLICATION NUMBER: US/09/671,317
CURRENT FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US 09/536,178
PRIOR FILING DATE: 2000-03-23
PRIOR APPLICATION NUMBER: PCT/IB00/00403
PRIOR FILING DATE: 2000-03-24
PRIOR APPLICATION NUMBER: US 60/126,269
PRIOR FILING DATE: 1999-03-25
PRIOR APPLICATION NUMBER: US 60/131,961
PRIOR FILING DATE: 1999-04-30
NUMBER OF SEQ ID NOS: 977
SOFTWARE: Patent.pm
SEQ ID NO 14
LENGTH: 929
TYPE: DNA
ORGANISM: Homo Sapiens

```

```

FEATURE:
NAME/KEY: allele
LOCATION: 501
OTHER INFORMATION: 12-454-363 : polymorphic base A or G
NAME/KEY: misc_binding
LOCATION: 481..500
OTHER INFORMATION: 12-454-363. misl, potential
NAME/KEY: misc_binding
LOCATION: 502..521
OTHER INFORMATION: 12-454-363. mis2, potential complement
NAME/KEY: primer bind
LOCATION: 139..158
OTHER INFORMATION: upstream amplification primer
NAME/KEY: primer bind
LOCATION: 634..652
OTHER INFORMATION: downstream amplification primer, complement
NAME/KEY: misc_binding
LOCATION: 489..513
OTHER INFORMATION: 12-454-363 potential probe
NAME/KEY: misc feature
LOCATION: 674..679, 881..882, 892..893
OTHER INFORMATION: n=a, g, c or t
US-09-671-317-14

```

```

Query Match      1.6%; Score 39.2; DB 4; Length 929;
Best Local Similarity 51.9%; Pred. No. 0.52;
Matches 83; Conservative 0; Mismatches 77; Indels 0; Gaps 0;

```

```

Oy 2048 GGCAATGCCAGTACGATGTGTTAGGCAAGAAAGCAACCTGAGCCAGAGTCCCTCAAGTA 2107
Db 915 GCCATTCACCTCCGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 856

Oy 2108 TTCAAGCGATGAGAGCCAGTACGAGAAACCTGAGCCAGAGTCCCTCAAGTA 2167
Db 855 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 796

Oy 2168 GAGACTGAAAAAGTAGAGCCCACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2207
Db 795 GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 756

```

```

RESULT 42
US-09-134-001C-1191
Sequence 1191, Application US/09134001C
Patent No. 6380370

```

```

GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 1191
LENGTH: 1839
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-1191

```

```

Query Match      1.6%; Score 39; DB 4; Length 1839;
Best Local Similarity 48.8%; Pred. No. 0.84;
Matches 105; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

```

```

Oy 1513 AAATTCGAATTGATATCTGAGAGCAAGTTCGATTGCTCAATTAGCTGATTAAGTA 1572
Db 138 AAGTATGAGAGCTAAATAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 197

Oy 1573 TACAAGCTCAGATGTTACATTTTGTATGAAATGATATATATCATGATGATGAGAGATGC 1632

```

Db 198 TGAACATTATATGAAGCGCTTAAGTCTTGAAGACGAGTTAGTACAAAATTAGAAAAGCT 257
QY 1633 ATATGTAAAGCCTCATATATGAGCCATAGTACATGATTTGAAAGATAGCTTTCTGATTA 1692
Db 258 ATATGTATACGACATTTTAAACAGATCAAGTACTGCAAAATGATATATACCGGTTT 317
QY 1693 GGAAGAAAGTTGAGCTCAAGCCTATATCTAAGAAA 1727
Db 318 AGAAGCGCGTGCACATCACTGTTATTAATATATA 352

RESULT 43
US-08-360-606B-29
; Sequence 29, Application US/08360606B
; Patent No. 5919617
; GENERAL INFORMATION:
; APPLICANT: Jnanendra K. Bhattacharjee
; APPLICANT: Richard C. Garrud
; APPLICANT: Paul L. Skatrud
; APPLICANT: Robert P. Peery
; TITLE OF INVENTION: Methods and Reagents for
; TITLE OF INVENTION: Detecting Fungal Pathogens in a
; TITLE OF INVENTION: Biological Sample
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSER: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 S. Wacker Drive Suite 3200
; CITY: Chicago
; STATE: Illinois
; COUNTRY: U.S.A.
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MS Word 7.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/360,606B
; FILING DATE: December 21, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Berghoff, Paul H.
; REGISTRATION NUMBER: 30,243
; REFERENCE/DOCKET NUMBER: 94,319
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312)913-0001
; TELEFAX: (312)913-0002
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1856 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: *Saccharomyces cerevisiae*
US-08-360-606B-29

Query Match 1.6%; Score 38.8; DB 2; Length 1856;
Best Local Similarity 50.3%; Pred. No. 0.96;
Matches 94; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 1276 AACTGCTAAAGAAAGAAATGTCCTCGTACCAAGATTTTATGATTAAGCATATTA 1335
Db 612 AACTCTCTTAAGAGAGATATATTTTGGTTTAAAGATTAACCTGAAAAGAAACTTT 671
QY 1336 TCTGTTAACTGAGCTCAATAGCCTTGTGNAATAAGGTCGTAATTCGATTTTCA 1395
Db 622 CCCATTATATTCGAAATATTCATTTCTCATTTGTTATAAGATCAAGCTGGTTGGCA 731
QY 1396 AGCCTTAACAATATTATTAAGACGCTTGATGATGATGATGATTAATTAAGAAAATTGGT 1455
Db 732 AGATGTTTAAAGAAAGATTCACCAAGGTAATGATATATTAATGATTAATTAATTTTGA 791

QY 1456 AGATGAT 1462
Db 792 AATGAT 798

RESULT 44
US-08-861-464-13/c
; Sequence 13, Application US/08861464
; Patent No. 5874210
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard P.
; APPLICANT: Austriaco Jr., Nicanor
; APPLICANT: Kennedy, Brian
; TITLE OF INVENTION: Genes Determining Cellular Senescence
; TITLE OF INVENTION: in Yeast
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/861,464
; FILING DATE: 22-MAY-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/396,001
; FILING DATE: 28-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/09351
; FILING DATE: 15-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/107,408
; FILING DATE: 16-AUG-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-6408A22
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 781-861-6240
; TELEFAX: 781-861-9540
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2150 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 563..1987
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 563..1987
US-08-861-464-13

Query Match 1.6%; Score 38.8; DB 2; Length 2150;
Best Local Similarity 51.1%; Pred. No. 1;
Matches 91; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

QY 2090 GAAGTCCAAATTAAGATTTCAAGCGGATGAAGACCCAGTAGAGAAAACCTGCTAG 2149
Db 1152 GAAGAGAAAG 1093
QY 2150 CGAAGAGCCTCAAGTAGAGACTGAAAAGTAAGTAAAGCCCAACTCAAGAGAGCAAGTT 2209

Db 1092 GTAGAAGTAGATGTGAAGTAGAAGACTAGATTAGTGGCAGAGAAGCACTAGAT 1033
Qy 2210 TTGCTTGGCAAGTAGAAGTAGAAGACTAGATTAGTGGCAGAGAAGCACTAGAT 2267
Db 1032 TGGGTTGGGGAAGAGTTGACACGACAGTAGATTGTTGAGCAGCTTTCTTCACTAGATG 975

RESULT 45

US-08-396-001-13/C
; Sequence 13, Application US/08396001
; Patent No. 5919618
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard P.
; APPLICANT: Austriaco Jr., Nicanor
; APPLICANT: Claus, James
; APPLICANT: Cole, Francesca
; APPLICANT: Kennedy, Brian
; TITLE OF INVENTION: Genes Determining Cellular Senescence in
; TITLE OF INVENTION: Yeast
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/396,001
; APPLICATION NUMBER: US/08/396,001
; FILING DATE: 28-FEB-1995
; CLASSIFICATION: .435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-6408A2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2150 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 563..1987
US-08-396-001-13

Query Match 1.6%; Score 38.8; DB 2; Length 2150;

Best Local Similarity 51.1%; Pred. No. 1; Mismatches 87; Indels 0; Gaps 0;

Matches 91; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 2090 GAAGATCCAAATATAGAACTTCAAGCGGATGAAGAGCCAGTAGAGAGAAACCTCTGAG 2149
Db 1152 GAAAGAGAAAG 1093
Qy 2150 CCAGAAGTCCCTCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGCAGAACTT 2209
Db 1092 GTAGAAGTAGATGTGAAGTAGAAGACTAGATTAGTGGCAGAGAAGCACTAGAT 1033
Qy 2210 TTGCTTGGCAAGTAGAAGTAGAAGACTAGATTAGTGGCAGAGAAGCACTAGAT 2267
Db 1032 TGGGTTGGGGAAGAGTTGACACGACAGTAGATTGTTGAGCAGCTTTCTTCACTAGATG 975

RESULT 46

US-09-323-433A-13/C
; Sequence 13, Application US/09323433A
; Patent No. 6218512
; GENERAL INFORMATION:
; APPLICANT: Guarente, Leonard P.
; APPLICANT: Austriaco Jr., Nicanor
; APPLICANT: Claus, James J.
; APPLICANT: Cole, Francesca
; APPLICANT: Kennedy, Brian
; TITLE OF INVENTION: GENES DETERMINING CELLULAR SENESENCE IN
; TITLE OF INVENTION: YEAST
; FILE REFERENCE: 0050.1491-003
; CURRENT APPLICATION NUMBER: US/09/323,433A
; PRIOR FILING DATE: 1999-06-01
; PRIOR APPLICATION NUMBER: US 08/396,001
; PRIOR FILING DATE: 1995-02-28
; PRIOR APPLICATION NUMBER: PCT/US94/09351
; PRIOR FILING DATE: 1994-08-15
; PRIOR APPLICATION NUMBER: US 08/107,408
; PRIOR FILING DATE: 1993-08-16
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 13
; LENGTH: 2150
; TYPE: DNA
; ORGANISM: Saccharomyces cerevisiae
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (563)...(1987)
; OTHER INFORMATION: SAG1
US-09-323-433A-13

Query Match 1.6%; Score 38.8; DB 3; Length 2150;

Best Local Similarity 51.1%; Pred. No. 1; Mismatches 87; Indels 0; Gaps 0;

Matches 91; Conservative 0; Mismatches 87; Indels 0; Gaps 0;

Qy 2090 GAAGATCCAAATATAGAACTTCAAGCGGATGAAGAGCCAGTAGAGAGAAACCTCTGAG 2149
Db 1152 GAAAGAGAAAG 1093
Qy 2150 CCAGAAGTCCCTCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGCAGAACTT 2209
Db 1092 GTAGAAGTAGATGTGAAGTAGAAGACTAGATTAGTGGCAGAGAAGCACTAGAT 1033
Qy 2210 TTGCTTGGCAAGTAGAAGTAGAAGACTAGATTAGTGGCAGAGAAGCACTAGAT 2267
Db 1032 TGGGTTGGGGAAGAGTTGACACGACAGTAGATTGTTGAGCAGCTTTCTTCACTAGATG 975

RESULT 47

US-08-119-125A-2
; Sequence 2, Application US/08119125A
; Patent No. 5610011
; GENERAL INFORMATION:
; APPLICANT: SMITH, Hilda Elizabeth
; APPLICANT: VECHT, Uri
; TITLE OF INVENTION: DNA Sequences which code for Virulence
; TITLE OF INVENTION: Characteristics of Streptococcus suis and parts thereof, polype
; TITLE OF INVENTION: antibodies derived therefrom and the use thereof for the diagn
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Central Diagenesekundig Instituut
; STREET: Edelhertweg 15
; CITY: PH Lelystad
; STATE:
; COUNTRY: The Netherlands
; ZIP: NL-8219
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS v.6.0


```
Qy 151 CATCAAGATTAACAGACCAAGGCTATGTCATTACATGCGCACCACTATCATTTATTACAA 210
Db 5971 TAAGATGAAGAAGCTTAATTAATGATTAACGAAATTTAATGATGAATTAATTAATGTTAA 6030
Qy 211 TGGTAAAGTCTCTTATGACGCTATCATCACTAAGAAATTTACTCATGAAAGATCCAACTA 270
Db 6031 TATGAATGTTCTTAATTAATTAATGCTAGTGCATTTGTAATGGAAGGATGAATAATGA 6090
Qy 271 TAAGCTAAAGATGAGATATTGTTAATGAGGTCAAGGGTGGATA 315
Db 6091 TAATTAGAAATGATGTTGTTGAAATAATATGATCATACATA 6135

RESULT 49
US-09-150-741-1
; Sequence 1, Application US/09150741
; Patent No. 6183996
; GENERAL INFORMATION:
; APPLICANT: Stewart et al.
; TITLE OF INVENTION: Nucleotide Sequence Encoding Carbamoyl Phosphate
; Patent No. 6183996
; TITLE OF INVENTION: Synthetase II
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/09/150,741
; CURRENT FILING DATE: 1998-09-10
; EARLIER APPLICATION NUMBER: PL6380
; EARLIER FILING DATE: 1992-12-16
; EARLIER APPLICATION NUMBER: AU93/00617
; EARLIER FILING DATE: 1993-12-02
; EARLIER APPLICATION NUMBER: 08/446,855
; EARLIER FILING DATE: 1995-07-06
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 8920
; TYPE: DNA
; ORGANISM: Plasmodium falciparum
US-09-150-741-1

Query Match 1.6%; Score 38.6; DB 3; Length 8920;
Best Local Similarity 46.0%; Pred. No. 2.4;
Matches 131; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

Qy 31 AACGGTTAGGAAATAATCGTTCTTATATAGATGAATAAACAAGCGACCAAAATAC 90
Db 5851 AATAGGAAATATAATAATAGATATGATTTATCTAAGGAAAAAGTATATCTAATA 5910
Qy 91 GGAGATTGACTCCTGATGAGGTAGCAAGCGTGAAGAAATCAATGCTGAGCAAAATCGT 150
Db 5911 AAACCTGTAATCTTATTTATGTTGATTCGCTATATAATAATGAATCAAAATTTAA 5970
Qy 151 CATCAAGTAAACAGACCAAGGCTATGTCATTCACTGACGCCACCTATCATTTATTACAA 210
Db 5971 TAAGATGAAGAAGTAAATAGATTAACGAAATTTAATGATGAATTAATTAATGTTAA 6030
Qy 211 TGGTAAAGTCTCTTATGACGCTATCATCACTAAGAAATTTACTCATGAAAGATCCAACTA 270
Db 6031 TATGAATGTTCTTAATTAATTAATGCTAGTGCATTTGTTAAATGGAAGGATGAATAATGA 6090
Qy 271 TAAGCTAAAGATGAGATATTGTTAATGAGGTCAAGGGTGGATA 315
Db 6091 TAATTAGAAATGATGTTGTTGAAATAATATGATCATACATA 6135

RESULT 50
US-08-169-927-1
; Sequence 1, Application US/08169927
; Patent No. 5783441
; GENERAL INFORMATION:
; APPLICANT: Carl, Mitchell
; APPLICANT: Dobson, Michael E.
; APPLICANT: Ching, Wei Mei
; APPLICANT: Dasch, Gregory A
```

```
; TITLE OF INVENTION: Gene and Protein Applicable to the
; TITLE OF INVENTION: Preparation of Vaccines for Rickettsia prowazekii and
; TITLE OF INVENTION: Rickettsia typhi and the Detection of Both
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Counsel, Naval Medical R & D Command
; STREET: Bldg. 1, T-12, 8901 Wisconsin Ave.
; CITY: Bethesda
; STATE: MD
; COUNTRY: USA
; ZIP: 20889-5606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/169,927
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/742,128
; FILING DATE: 08/09/91
; ATTORNEY/AGENT INFORMATION:
; NAME: Spevack, A. David
; REGISTRATION NUMBER: 24,743
; REFERENCE/DOCKET NUMBER: 75,976
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 295-6759
; TELEFAX: (301) 295-1022
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5319 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; ORIGINAL SOURCE:
; ORGANISM: Rickettsia prowazekii
; STRAIN: Breinl
; FEATURE:
; NAME/KEY: -35_signal
; LOCATION: 340..345
; FEATURE:
; NAME/KEY: -10_signal
; LOCATION: 363..368
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 391..5226
; FEATURE:
; NAME/KEY: RBS
; LOCATION: 379..386
; FEATURE:
; NAME/KEY: stem_loop
; LOCATION: 5270..5306
; PUBLICATION INFORMATION:
; AUTHORS: Carl, M.
; AUTHORS: Dobson, M. E.
; AUTHORS: Ching, W. M.
; AUTHORS: Dasch, G. A.
; TITLE: Characterization of the gene encoding the
; TITLE: protective S-layer protein of Rickettsia
; TITLE: prowazekii; presence of a truncated identical
; TITLE: homolog in rickettsia typhi
; JOURNAL: Proc. Natl. Acad. Sci. U.S.A.
; DATE: 1990
; RELEVANT RESIDUES IN SEQ ID NO: 1: FROM 1 TO 5319
US-08-169-927-1

Query Match 1.6%; Score 38.4; DB 1; Length 5319;
```

Best Local Similarity 48.6%; Pred.No. 2.1; Matches 105; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

Oy	1581	CAGATGGTTACATTTTGGATGAACATGATATATCAGTGTAGAGAGATGCATATGTAA	1640
Db	1067	CAGGTGTAATCTATTATTTAATGAAATGACGGTACTGTAAATTAGTATTAGTCA	1126
Oy	1641	CGCCTCATATGGGCCATAGTCACTGGATTGAAAGATAGCCTTCTGATAAGAAAAAG	1700
Db	1127	GTAAGATGTGCTGCTACCGAATTTAATGTACAGAACTTTAGGTGTAATCTAAAG	1186
Oy	1701	TTGCAGCTCAAGCCTATTAATAAGAAAAAGTATCCTCACTCCATCTCCAGACGAGATG	1760
Db	1187	GTATTATTGAATTGAACACTGCAGCAGTAGCTGTAACTTACTCTCTTGAGGGTGTG	1246
Oy	1761	TTAAGCAAAATCCAACTGGAGATAGTGCAGCAGCTA	1796
Db	1247	CTAATGCAGTAATAGGTACAGATAAATGGACAGTA	1282

Search completed: November 14, 2003, 01:12:36
Job time : 179 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM nucleic - nucleic search, using sw model.

Run on: November 13, 2003, 23:47:23 ; Search time 758 Seconds
(without alignments)
10300.449 Million cell updates/sec

Title: US-09-765-271-55

Perfect score: 2389

Sequence: 1 TTCTTACGAGTTGGACTGT.....TAAGTACGAAAAATAAAC 2389

Scoring table: IDENTITY NUC
Gapop 10.0, Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : Published Applications, NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2389	100.0	2389	9	US-09-765-272-55
2	1374.2	57.5	2523	11	US-09-884-465A-3
3	1374.2	57.5	2647	11	US-09-884-465A-4
4	1003.8	42.0	2481	11	US-09-769-787-206
5	987.6	41.3	2290	9	US-09-765-272-65
6	980.8	41.1	2639	11	US-09-884-465A-5
7	385.4	16.1	1382	9	US-09-765-272-181
8	385.4	16.1	1455	11	US-09-769-787-246
9	385.4	16.1	1455	12	US-09-769-744A-23
10	385.4	16.1	3120	11	US-09-884-465A-1
11	385.4	16.1	5048	11	US-09-884-465A-2
12	347	14.5	2528	11	US-09-884-465A-9
13	211.4	8.8	1166	12	US-09-769-736-23
14	211.4	8.8	2469	12	US-09-769-736-17
15	209.8	8.8	5215	11	US-09-252-088-13
16	201.2	8.4	819	11	US-09-884-465A-257

17	117.4	4.9	1455	12	US-09-769-736-71	Sequence 71, Appl
18	109	4.6	841	9	US-09-452-599-34	Sequence 34, Appl
19	109	4.6	841	12	US-10-121-120-34	Sequence 34, Appl
20	51.2	2.1	1168	12	US-10-017-161-2179	Sequence 2179, Ap
21	50	2.1	573	12	US-10-029-386-25433	Sequence 25433, A
22	49.8	2.1	1559	9	US-09-864-761-4012	Sequence 4012, Ap
23	46.6	2.0	506	12	US-10-029-386-20619	Sequence 20619, A
24	45.4	1.9	58985	11	US-09-901-152-3	Sequence 3, Appl
25	45.4	1.9	143601	12	US-09-855-824-3	Sequence 3, Appl
26	44.6	1.9	423	9	US-09-864-761-19355	Sequence 18355, A
27	44.6	1.9	487	9	US-09-864-761-1597	Sequence 1597, A
28	44.6	1.9	3673778	12	US-10-312-841-2	Sequence 2, Appl
29	43.6	1.8	766	9	US-09-864-761-19608	Sequence 19608, A
30	43.6	1.8	1944	9	US-09-864-761-2825	Sequence 2825, A
31	43	1.8	538	12	US-10-029-386-20406	Sequence 20406, A
32	42.6	1.8	785	12	US-10-029-386-22627	Sequence 22627, A
33	42.4	1.8	660	9	US-09-864-761-19488	Sequence 19488, A
34	42.4	1.8	962	9	US-09-864-761-2772	Sequence 2772, Ap
35	42.4	1.8	4047	9	US-09-815-242-4843	Sequence 4843, Ap
36	42.4	1.8	4050	9	US-09-815-242-9039	Sequence 9039, Ap
37	42.4	1.8	4358	8	US-08-781-986A-454	Sequence 454, App
38	42.2	1.8	635	12	US-10-027-632-26927	Sequence 26927, A
39	42.2	1.8	635	13	US-10-027-632-26927	Sequence 26927, A
40	42	1.8	305	9	US-09-864-761-19262	Sequence 19262, A
41	42	1.8	496	9	US-09-864-761-2534	Sequence 2534, Ap
42	42	1.8	3532	12	US-10-017-161-1913	Sequence 1913, Ap
43	42	1.8	528	12	US-10-029-386-5350	Sequence 5350, Ap
44	41.8	1.7	1010	12	US-10-029-386-24901	Sequence 24901, A
45	41.8	1.7	3673778	12	US-10-312-841-1	Sequence 1, Appl
46	41.6	1.7	413	12	US-10-029-386-14986	Sequence 14986, A
47	41.6	1.7	571	12	US-10-029-386-1284	Sequence 1284, Ap
48	41.6	1.7	611	10	US-09-864-864-233	Sequence 233, App
49	41.6	1.7	1845	10	US-09-864-864-313	Sequence 313, App
50	41.6	1.7	1860	10	US-09-919-487-50	Sequence 50, Appl
51	41.4	1.7	5361	10	US-09-742-036-2	Sequence 2, Appl
52	41.4	1.7	6152	10	US-09-742-036-1	Sequence 1, Appl
53	41.2	1.7	696	10	US-09-922-261-193	Sequence 193, App
54	41.2	1.7	699	10	US-09-922-261-191	Sequence 191, App
55	41.2	1.7	700	13	US-10-040-739-1385	Sequence 1385, Ap
56	41.2	1.7	717	10	US-09-922-261-189	Sequence 189, App
57	41.2	1.7	774	10	US-09-922-261-187	Sequence 187, App
58	41.2	1.7	819	10	US-09-922-261-185	Sequence 185, App
59	41.2	1.7	1669	10	US-09-922-261-184	Sequence 184, App
60	41	1.7	3996	14	US-10-087-464-42	Sequence 42, Appl
61	41	1.7	14066	12	US-10-349-680-149	Sequence 149, App
62	40.8	1.7	286	12	US-10-029-386-18943	Sequence 18943, A
63	40.8	1.7	577	12	US-10-029-386-5187	Sequence 5187, A
64	40.8	1.7	579	12	US-10-029-386-25309	Sequence 25309, A
65	40.8	1.7	2000	10	US-09-938-882A-4612	Sequence 4612, Ap
66	40.8	1.7	2013	12	US-10-032-585-6568	Sequence 6568, Ap
67	40.6	1.7	276	9	US-09-864-761-20595	Sequence 20595, A
68	40.4	1.7	1372	12	US-10-017-161-2245	Sequence 2245, Ap
69	40.4	1.7	7331	12	US-10-311-455-2158	Sequence 2158, Ap
70	40.2	1.7	475	9	US-09-864-761-1619	Sequence 1619, Ap
71	40.2	1.7	536	12	US-10-029-386-19974	Sequence 19974, A
72	40.2	1.7	599	12	US-10-029-386-6243	Sequence 6243, Ap
73	40	1.7	4316	10	US-09-880-107-3713	Sequence 3713, Ap
74	40	1.7	16239	14	US-10-175-553-64	Sequence 3713, Ap
75	39.8	1.7	8293	12	US-10-311-455-32	Sequence 64, Appl
76	39.6	1.7	1308	12	US-10-349-680-100	Sequence 32, Appl
77	39.6	1.7	1458	9	US-09-815-242-4757	Sequence 100, App
78	39.6	1.7	1484	14	US-10-234-432-12	Sequence 4757, Ap
79	39.6	1.7	1599	9	US-09-815-242-8824	Sequence 12, Appl
80	39.6	1.7	1768	14	US-10-234-432-48	Sequence 8824, Ap
81	39.6	1.7	1855	14	US-10-234-432-47	Sequence 48, Appl
82	39.6	1.7	1855	14	US-10-234-432-52	Sequence 47, Appl
83	39.6	1.7	2856	14	US-10-234-432-50	Sequence 52, Appl
84	39.6	1.7	2875	14	US-10-234-432-50	Sequence 50, Appl
85	39.6	1.7	18355	8	US-08-781-986A-67	Sequence 67, Appl
86	39.6	1.7	143899	11	US-09-972-586-15	Sequence 15, Appl
87	39.6	1.7	3673778	12	US-10-312-841-2	Sequence 15, Appl
88	39.4	1.6	315	9	US-09-864-761-21723	Sequence 21723, A
89	39.4	1.6	40862	12	US-10-311-455-2046	Sequence 2046, Ap

C 90 39.2 1.6 462 9 US-09-864-761-3829 Sequence 3829, Ap
C 91 39.2 1.6 1074 9 US-09-861-451A-49 Sequence 49, Appl
C 92 39.2 1.6 1495 12 US-10-032-585-6091 Sequence 6091, Ap
C 93 39.2 1.6 6944 14 US-10-172-086-111 Sequence 111, App
C 94 39 1.6 457 9 US-09-864-761-5436 Sequence 5436, Ap
C 95 39 1.6 2050 14 US-10-086-510-1 Sequence 1, Appl
C 96 38.8 1.6 277 9 US-09-294-093B-1831 Sequence 1831, Ap
C 97 38.8 1.6 505 10 US-09-917-800A-314 Sequence 314, App
C 98 38.8 1.6 510 9 US-09-864-761-18737 Sequence 18737, A
C 99 38.6 1.6 2150 9 US-09-826-752-13 Sequence 13, Appl
C 100 38.6 1.6 511 12 US-10-027-632-302235 Sequence 302235,

ALIGNMENTS

RESULT 1
US-09-765-272-55
Sequence 55, Application US/09765272
Patent No. US20020061545A1

GENERAL INFORMATION:
APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 55:
SEQUENCE CHARACTERISTICS:
LENGTH: 2389 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 55:
US-09-765-272-55

Query Match 100.0%; Score 2388; DB 9; Length 2389;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTCTTACGAGTTGGACGTGTATCAAGCTAGAAACGGTTAAAGAAATAATCGTGTCTTCTA 60
DB 1 TTCTTACGAGTTGGACGTGTATCAAGCTAGAAACGGTTAAAGAAATAATCGTGTCTTCTA 60
QY 1 TTCTTACGAGTTGGACGTGTATCAAGCTAGAAACGGTTAAAGAAATAATCGTGTCTTCTA 120
DB 1 TTCTTACGAGTTGGACGTGTATCAAGCTAGAAACGGTTAAAGAAATAATCGTGTCTTCTA 120
QY 121 GCGTGAAGAAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGTCAC 180
DB 121 GCGTGAAGAAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGTCAC 180

DB 121 GCGTGAAGAAATCAATGCTGAGCAAAATCGTCAATCAAGATAACAGACCAAGGCTATGTCAC 180
QY 181 TTTCACATGCGACACCATATCATCTTATTAACAAATGTAAGTTCCTATAGACGTATCATCAG 240
DB 181 TTTCACATGCGACACCATATCATCTTATTAACAAATGTAAGTTCCTATAGACGTATCATCAG 240
QY 241 TGAAGAAATTAATCAATGAAAGATCCAAATTAATAGCTAAAGATAGATATGTTAAATGA 300
DB 241 TGAAGAAATTAATCAATGAAAGATCCAAATTAATAGCTAAAGATAGATATGTTAAATGA 300
QY 301 GGTCAAGGTTGATATGTTATCAAGGTATGAGAAATTAATGATTTAATCTTAAGATGC 360
DB 301 GGTCAAGGTTGATATGTTATCAAGGTATGAGAAATTAATGATTTAATCTTAAGATGC 360
QY 361 TGCCACGCGGATATACCTCGTACAAAGGAAATCAATGACAAAGGATAG 420
DB 361 TGCCACGCGGATATACCTCGTACAAAGGAAATCAATGACAAAGGATAG 420
QY 421 TCAACATGCTGAAAGTGAATCTCCAGAAACGATGCTGCTTGCCTTGACAGCTTCGCA 480
DB 421 TCAACATGCTGAAAGTGAATCTCCAGAAACGATGCTGCTTGCCTTGACAGCTTCGCA 480
QY 481 AGAAGCTATATCAACAGATGATGTTATATCTTTAATGCTTGAATATCATAGAGATAC 540
DB 481 AGAAGCTATATCAACAGATGATGTTATATCTTTAATGCTTGAATATCATAGAGATAC 540
QY 541 TGGTGAATGCTTATATGCTTCTCCTCATGAGATCATTAACATTAATCTTAAGATGTT 600
DB 541 TGGTGAATGCTTATATGCTTCTCCTCATGAGATCATTAACATTAATCTTAAGATGTT 600
QY 601 ATCAGTACGAGATGCTGCTGTCGAGAAAGCTTCTATCTGTCGAGAAATCTGTCAA 660
DB 601 ATCAGTACGAGATGCTGCTGTCGAGAAAGCTTCTATCTGTCGAGAAATCTGTCAA 660
QY 661 TTCAAGAACTATATGCTGTCGAGAAAGCTTCTATCTGTCGAGAAATCTGTCAA 720
DB 661 TTCAAGAACTATATGCTGTCGAGAAAGCTTCTATCTGTCGAGAAATCTGTCAA 720
QY 721 TGTAGAAATTCAGAACTAATTAATCAACAGAAACGAAACCTAAGATCA 780
DB 721 TGTAGAAATTCAGAACTAATTAATCAACAGAAACGAAACCTAAGATCA 780
QY 781 AGCAAGTCAAAAGTATATGATATGCTTCTTGAACAGCTCTCAAACTGCTTTGAG 840
DB 781 AGCAAGTCAAAAGTATATGATATGCTTCTTGAACAGCTCTCAAACTGCTTTGAG 840
QY 841 TCAACGACATGTATGATATGATGCTTCTTGAACAGCTCTCAAACTGCTTTGAG 900
DB 841 TCAACGACATGTATGATATGATGCTTCTTGAACAGCTCTCAAACTGCTTTGAG 900
QY 901 AGCTAGAGGTTGCTGACGTCAGACGAGATCAATTAACACTTAATCTCTCAAT 960
DB 901 AGCTAGAGGTTGCTGACGTCAGACGAGATCAATTAACACTTAATCTCTCAAT 960
QY 961 GTCTGAATTTGGAAGAAAGAAATCGCTGTAATTAATCCCTTCTGTTATGCTTCAATG 1020
DB 961 GTCTGAATTTGGAAGAAAGAAATCGCTGTAATTAATCCCTTCTGTTATGCTTCAATG 1020
QY 1021 GGTACAGATTTCAAGGCTCAGAACCAAGCTCAACCGACTCGGAACCTAGTCCAGG 1080
DB 1021 GGTACAGATTTCAAGGCTCAGAACCAAGCTCAACCGACTCGGAACCTAGTCCAGG 1080
QY 1081 CCGGCAACCTGCAACCAATCTTAATTAATGACTCAATCTTCTTGTGTTAGTCAAGTGGT 1140
DB 1081 CCGGCAACCTGCAACCAATCTTAATTAATGACTCAATCTTCTTGTGTTAGTCAAGTGGT 1140
QY 1141 ACAGAAAGTTGGGAGATATGATTTCAAGAAAGGCAATCTCTGTTATGCTTTGC 1200
DB 1141 ACAGAAAGTTGGGAGATATGATTTCAAGAAAGGCAATCTCTGTTATGCTTTGC 1200
QY 1201 GAAAGATTTACATCTGAAAATCTTTAAATAATCTTGAAGCAAGTTATCAAAACAAAGAG 1260
DB 1201 GAAAGATTTACATCTGAAAATCTTTAAATAATCTTGAAGCAAGTTATCAAAACAAAGAG 1260

QY 1261 TGTTCACACACTTAACTGCTAATAAAAAAGTGTGCTCTGTCGACCAAGATTTTA 1320
Db 1261 TGTTCACACACTTAACTGCTAATAAAAAAGTGTGCTCTGTCGACCAAGATTTTA 1320
QY 1321 TGAATAAGCATATATCTGTAACTGAGGCTCATAAAGCTTTGTTGNAATAAAGGTCG 1380
Db 1321 TGAATAAGCATATATCTGTAACTGAGGCTCATAAAGCTTTGTTGNAATAAAGGTCG 1380
QY 1381 TAATTCGATTTTCCAAGCCTTGAACAATAATATAGAACGCTTAAGTGAATCGACTAA 1440
Db 1381 TAATTCGATTTTCCAAGCCTTGAACAATAATATAGAACGCTTAAGTGAATCGACTAA 1440
QY 1441 TAAAGAAAAATGATGATGATTTATGTCATTCAGACCAATTAACCATCCAGACG 1500
Db 1441 TAAAGAAAAATGATGATGATTTATGTCATTCAGACCAATTAACCATCCAGACG 1500
QY 1501 ACTTGGCAACCAAAATTTCTCAAATTGAGTATCTGAAGACGAAGTTGTAATTCCTCAAT 1560
Db 1501 ACTTGGCAACCAAAATTTCTCAAATTGAGTATCTGAAGACGAAGTTGTAATTCCTCAAT 1560
QY 1561 AGCTGATTAAGTATACAAACGTCAGATGCTTATCTTTGATGAACATGATATATCAAGTGA 1620
Db 1561 AGCTGATTAAGTATACAAACGTCAGATGCTTATCTTTGATGAACATGATATATCAAGTGA 1620
QY 1621 TGAAGAGATGATATATGTAACGCTCATATGAGGCAATGATCACTGATTTGAAAAAGTAG 1680
Db 1621 TGAAGAGATGATATATGTAACGCTCATATGAGGCAATGATCACTGATTTGAAAAAGTAG 1680
QY 1681 CCTTTCTGATAGAAAAAGTTGCAAGCTCAAGCTTACTTAAAGAAAAAGTATCTTACC 1740
Db 1681 CCTTTCTGATAGAAAAAGTTGCAAGCTCAAGCTTACTTAAAGAAAAAGTATCTTACC 1740
QY 1741 TCCATCTCAGAGCAGATGTTAAAGCAATCCAACTGAGATAGTGCACACTTATTA 1800
Db 1741 TCCATCTCAGAGCAGATGTTAAAGCAATCCAACTGAGATAGTGCACACTTATTA 1800
QY 1801 CAATCGTGTGAAAAAGGAAAAAGCAATTCACCTGTTCCAGCTTCCATATATGTTGAGCA 1860
Db 1801 CAATCGTGTGAAAAAGGAAAAAGCAATTCACCTGTTCCAGCTTCCATATATGTTGAGCA 1860
QY 1861 TACAGTTGAGTTAAAAAGGTAATTTGATTTATCTCATTAAGATCATTAACATATAT 1920
Db 1861 TACAGTTGAGTTAAAAAGGTAATTTGATTTATCTCATTAAGATCATTAACATATAT 1920
QY 1921 TAAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980
Db 1921 TAAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980
QY 1981 TTTGTTTGGGACGATTAAGTACTAGTGAACAACCTGACGAAGTCCACATTTCTAATGA 2040
Db 1981 TTTGTTTGGGACGATTAAGTACTAGTGAACAACCTGACGAAGTCCACATTTCTAATGA 2040
QY 2041 TGGATGGGGCAATGCCAGTGAAGCATGTGTAGGCAAGAAAGACACAGTGAATCCAAA 2100
Db 2041 TGGATGGGGCAATGCCAGTGAAGCATGTGTAGGCAAGAAAGACACAGTGAATCCAAA 2100
QY 2101 TAAAGAACTTCAAGCGGATGAAGACGATGAGGAAACCTGCTGAGCCAGAAAGTCCC 2160
Db 2101 TAAAGAACTTCAAGCGGATGAAGACGATGAGGAAACCTGCTGAGCCAGAAAGTCCC 2160
QY 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAAAGCAAGTTTGTTCGAA 2220
Db 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAAAGCAAGTTTGTTCGAA 2220
QY 2221 AGTAAAGGATTTCTAGTCTGAAGCAATGACAGAAACTCTAGTGTGTTTACGAATTA 2280
Db 2221 AGTAAAGGATTTCTAGTCTGAAGCAATGACAGAAACTCTAGTGTGTTTACGAATTA 2280
QY 2281 TTTGACTCTTCAAAATTAATGATTAACAATAGTATCAATGAGCAAGAAAGCAAAATTA 2340
Db 2281 TTTGACTCTTCAAAATTAATGATTAACAATAGTATCAATGAGCAAGAAAGCAAAATTA 2340

QY 2341 GTTGTAAAAAGAGTAATCTTTCATCTGTAAAGTAAAGAAAAATTAAC 2389
Db 2341 GTTGTAAAAAGAGTAATCTTTCATCTGTAAAGTAAAGAAAAATTAAC 2389
RESULT 2
US-09-884-465A-3
; Sequence 3, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josef
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 2523
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-3

Query Match 57.5%; Score 1374.2; DB 11; Length 2523;
Best Local Similarity 73.9%; Pred. No. 0;
Matches 1819; Conservative 0; Mismatches 539; Indels 102; Gaps 2;

QY 1 TTTCTACGAGTTGGGACTGTATCAAGCTTAAGAAAGGTTAAGAAATATCGTGTTCCTTA 60
Db 60 TGTCTATGAACTAGAGTTTGCATCAAGCTCAAACTGTAAAAAGAAATATCGTGTTCCTTA 119
QY 61 TATAGATGAAAAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAG 120
Db 120 TATAGATGAAAAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAG 179
QY 121 GCGTGAAGGAAATCAATGCTGAGCAAAATCGTCAATCAAGATTAAGACCAAGGCTATGTCAC 180
Db 180 GCGTGAAGGAAATCAAGCGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGGAAAGG 239
QY 181 TTCACATGGGCAACCACTATCATTTATTAAGTGAAGGTTCTTATGACGTATCATGAC 240
Db 240 CTCCTATGGAGACCAATTAATCATTAATGATGCAAGGTTCCCTTATGATGCCATCATGAC 299
QY 241 TGAAGAAATTAATCTGAAAGATCCAAACTATAAGCTAAAAAGATGAGGATATGTTAATGA 300
Db 300 TGAAGAGCTCCATGAAAGATCCGAATTAATCAAGTGAAGATTCAGACATTTGTCATATGA 359
QY 301 GGTCAAGGTGATATGTTATCAAGGTAGTGAAGAAATATGTTTACCTTAAGGATGC 360
Db 360 AATCAAGGTGATGTTATGATTAAGTGAAGGAAATATGTTTATTAAGGATGC 419
QY 361 TGCACAGGGGATTAAGTCCGTACAAAGAGGAAATCAATGACAAAGAAAGGAAAGGAAAG 420
Db 420 AGCTCATGGGATTAATGTCGTACAAAGAGGAAATCAATGCGGCAAGAAAGGAAAGGAAAG 479
QY 421 TCAACATCTGTAAGGTGAACTCCAAAGAAAGATGATGCTGTGCTTGGCACGTTCCGA 480
Db 480 TCAGCATCTGAAGGAGGACTTCAGCAAGCATGATGAGGTTGACCTTTGCAAGTTCAACA 539
QY 481 AGGAGCTATACATCAAGATGATGTTATATCTTTATGCTTCGATATCATGAGAGATAC 540
Db 540 GGGAGCTTACACACAGATGATGTTATATCTTTATGCTTCGATATCATGAGAGATAC 599
QY 541 TGTGATGCTTATATGCTTCTCATGAGATCATTAACATTAATTTCTTAAGATGAGTT 600
Db 600 GGGGATGCTTATATGCTTCTCATGAGATCATTAACATTAATTTCTTAAGATGAGTT 659

601 ATCAGTCGCGAGTGGCTGCTGTCAGAGAGCCCTTCCTATGCTGCGAGGAAATCTGTCAA 660
 660 ATCAGTCGCGAGTGGCTGCTGTCAGAGAGCCCTTCCTATGCTGCGAGGAAATCTGTCAA 719
 661 TTCAAGAACTATGCGCCGCAAAATAGCGATTAACCTTCAGAGCAAACTGGGATCCTTC 720
 720 TTAAAGAACTATGCGCCGCAAAATAGCGATTAACCTTCAGAGCAAACTGGGATCCTTC 779
 721 TGTAGCAATCCAGAACTAACAATACTTAACCAAGCAACAACAGCAACCTTAACAGTCA 780
 780 TGTAGCAATCCAGAACTAACAATACTTAACCAAGCAACAACAGCAACCTTAACAGTCA 839
 781 AGCAAGTCAAAAGTAAATGACATGATAGTCTCTGAAAGAGCTCTCAAACTGCTTGAAG 840
 840 AGCAAGTCAAAAGTAAATGACATGATAGTCTCTGAAAGAGCTCTCAAACTGCTTGAAG 899
 841 TCAAGCAATGATAGATCTGATGCGCTTGTCTTTGATCCAGCAAAATCACAAGTCGAAC 900
 900 TCAAGCAATGATAGATCTGATGCGCTTGTCTTTGATCCAGCAAAATCACAAGTCGAAC 959
 901 AGCTAGAGTGTGTCAGTGCACACGAGATCATTAACATTCCTTACTCTCAAT 960
 960 CGCCAGAGGTGATGCTGCTCCCTCATGATTAACCATTAACATTCCTTACTCTCAAT 1019
 961 GTCTGAATTTGGAAGAGCAATGCTGATTAATTCCTGCTTGTATTCGTTCAAACTATG 1020
 1020 GTCTGAATTTGGAAGAGCAATGCTGATTAATTCCTGCTTGTATTCGTTCAAACTATG 1079
 1021 GGATACAGATTCAGAGCGAGAAACAACAGTCCACAGCACTCCGGAACCTAGTCAGAG 1080
 1080 GGATACAGATTCAGAGCGAGAAACAACAGTCCACAGCACTCCGGAACCTAGTCAGAG 1139
 1081 CCGGCACTCTGACCAAACTTTAAATAGACTCA-----ATTCTTCTTGGT 1128
 1140 TCCGCAACCTGACCAAACTCTCAACAGCTCCAGCAATCCAAATGATGAGAAATTTGT 1199
 1129 TAGTACGTGTGAGCAAGAGTTGGGAGAGATATGATTTGGAAGAAAGGAGATCTCG 1188
 1200 CAAGAAAGCTGTTCAAAAGTAGGCGATGTTATGCTTTGAGGAATGGAGTTTCTCG 1259
 1189 TTATGCTTTTGGGAAGATTTACATCTGAAACTGTAAATCTTGAAGCAAGTTATC 1248
 1260 TTATATCCAGCAAGAACTTTTACGAGAAACAACAGCAAGCAATTTGATGAACCTGCG 1319
 1249 AAAACAAAGAGTGTTCACACACTTTAATGCTTAAAGAAAGAAATGTTGCTCTCGTGA 1308
 1320 CAAGCAAGAAAGTGTTCATTAAGCTAAGAGCTAAGAAACTGACCTCCCATCTAGTGA 1379
 1309 CCAAGAAATTTTATGATAAGCATATATCTGTTAAGAGCTCAATTAAGAGCTTGTGN 1368
 1380 TCGAATAATTTTCAATTAAGAGCTTATGACTTACTACCAAGAAATTTCAAGAAATTTTCA 1439
 1369 AAATTAAGAGTGTATTTCTGATTTTCAAGCTTACCAAAATTTATGAAGAGCTTGAATGA 1428
 1440 TAAATTAAGTGTGACAGATTTGATTTTGAAGCTTTGATTAACCTGTTGGAACAGCTCAAGA 1499
 1429 TGAATCGATTAATAAGAAATTTGATGATGATTTATTTGGATTCCTAGCAACAATTAC 1488
 1500 TGTCTCAAGTATTAAGTCAAGTGTAGTATGATTTCTTCCGCTTCTTGAAGCTCCGATTCG 1559
 1489 CCATCCAGAGGAGCTTGGCAAAACAATTCCTCAATTAAGATTAAGTGAAGAGAGTTCG 1548
 1560 TCAATCCAGAGGAGCTTGGCAAAACAATTCCTCAATTAAGATTAAGTGAAGAGATTTCA 1619
 1549 TATTGCTCAATTAAGTGTATGATTAACAGCTCAGATGTTATCATTTTGAATGAACATGA 1608
 1620 AGTAGCCAAAGTTGGAGGCAAGTACACAAGAAAGAGGATTAATCTTTGATCTCGTGA 1679
 1609 TATAATCAGTATGAAGAGATGATATGTAAAGCTCATATGAGGCAATGAGCTAGTCTGAT 1668
 1680 TATAACCAAGTATGAGGAGGATGCTATGTAATCTCAATATGAGCCCAATGAGCACTGAT 1739

1669 TGAAGAAAGATAGCTTTCTGATTAAGAAAAAGTTGACAGTCAAGCTTATCTAATAAGAAAA 1728
 1740 TAAAAAGATAGTGTGTCTGAGCGTAGAGAGCGGACGCCAGCTTATGCTTAAAGAGAA 1799
 1729 AGGTATCTTACCTTCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGAGATAGTGC 1788
 1800 AGGTTTGAACCCCTCTCTGACAGACCATCAAGATTCAGAAATCTGAGGCAAAAGAGGC 1859
 1789 AGCAGCTATTTAACATGCTGTGAAGAGGAAAAAGAAATTCACATCTGCTGCACTTCAATA 1848
 1860 AGAAGCTATCTTAACACGCGTGAAGAGAGCTTAAGAGAGTGCACCTGATCTATGCTTTA 1919
 1849 TATGTTGAGAGATCAGTTGAGTTGAATTAAGAGTAAATTTGATTAATTCCTCATTAAGATCA 1908
 1920 CAATCTTCAATATCTGTAGAGTCAAAAAGAGTATTAATCACTCATTTATGACCA 1979
 1909 TTACCTATATATTAATTTGCTGTTGATGATGATCAATACAAAGTCCCAATGCTTA 1968
 1980 TTACCTATATCAATCAAAATTTGAGTGTGACGAAGGCTTTATGAGGCACTTAAGGGGTA 2039
 1969 TACCTTGAAGATTTGTTTGGAGATTAAGTACTAAGTAAAGACCTGACGAACGTC 2028
 2040 TACTCTTGAAGATCTTTTGGAGCTGTCAGATTAATGTAACAATCCAAAGCAAGTCC 2099
 2029 ACATTTAATGATGATGAGGCAATGCGCAGTGAAGATGTTAGCAAGAAAGACCAAG 2088
 2100 GCATTCAGATTAATGTTTGTGTAACCTAGGACCAATGTTCAAAAGAAACAAATTTGCTCA 2159
 2089 TGAAGATCCAAATTAAGAACTTCAAGCGATGAAGAG----- 2125
 2160 AGTGTATTAACCAATCAAAACGAAAAACAAGGAGAGAAACCTCAGACAGAAAAACCTGA 2219
 2126 ----- 2125
 2220 GGAAGAAACCTCTGAGAGAGAAACAACAAGGAGAAACAGAGTCTCCAAAAACAAC 2279
 2126 -----CCAGTAGAGAAACACCTGCTGAGCCAGAACTCCCTCAAGTAGAGACTGAAAA 2178
 2280 AGAGAAACCAAGAAAGATTAACAGAGATTCAGAGAAACCTTCAAGTCCAGACTGAAAA 2339
 2179 AGTAAAGCCCACTCAAGAGAGAGAGTGTGCTTGGCAAGTAAAGATTTCTAGTCT 2238
 2240 GTTGAAGAAAACTGAGAGAGGCTGAAGATTTACTTGAAGAAATCCAGAGATCCAAATTA 2399
 2239 GAAGCCAAAGCAACAGAAACTCTAGCTGTTTCAAGTAAATTTGACTCTCAATTTAT 2238
 2400 CAAGTCCAAATGCAAGAGAGCTTCACAGAGTTAAAAATTAATTTACTATTTGSCACCCA 2459
 2299 GGATTAACATATGATTCATGCGAGAGAGAGAAAAATTAAGTCTGTTTAAAGAGAGTAA 2358
 2460 GGAACAACATATCTTATTTAGGCAAGAGCTGAAAACTATTTGCTTTATTTAAAGAGAGTAA 2519

RESULT 3
 US-09-884-465A-4
 ; Sequence 4, Application US/09884465A
 ; Publication No. US2003007293A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shire Biochem, Inc.
 ; APPLICANT: Hamel, Josee
 ; APPLICANT: Brodeur, Bernard
 ; APPLICANT: Martin, Denis
 ; APPLICANT: Charland, Nathalie
 ; APPLICANT: Ouellet, Catherine
 ; TITLE OF INVENTION: Streptococcus Antigens
 ; FILE REFERENCE: 055190-0044
 ; CURRENT APPLICATION NUMBER: US/09/884,465A
 ; PRIORITY FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: 60/212,683
 ; NUMBER OF SEQ ID NOS: 384
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 4

LENGTH: 2647
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-884-465A-4

Query Match 57.5%; Score 1374.2; DB 11; Length 2647;
Best Local Similarity 73.9%; Pred. No. 0;
Matches 1819; Conservative 0; Mismatches 539; Indels 102; Gaps 2;

```
QY 1 TTCTTCSAGTGGGACGTGATCAAGCTGTAAGGTTAAGGAAATATCGTCTTCCTA 60
DB 104 TCGTTATGAACTAGGTTTGATCACTCAACTGTAAGAAATATCGTCTTCCTA 163
QY 61 TATAGATGAAAAACAAGCAGCAAAAAACGAGATTGACTCTGATGAGGTAGCAA 120
DB 164 TATAGATGAAAAACAAGCAGCAAAAAACGAGATTGACTCTGATGAGGTAGCAA 223
QY 121 GCGTGAAGATCAATGCTGAGCAAAATCGTCATCAAGATPAACAGCAAGCTATGTGAC 180
DB 224 GCGTGAAGATCAACGCCGACAAATCGTCATCAAGATPAACAGCTATGTGAC 283
QY 181 TTCACATGCGACCATCATCTATTATACAATGTTAGTTCCTTATGACGCTATCATGAC 240
DB 284 CTCTCATGAGACCATTTATCTATTAATGTCAGAGTCCCTTATGATGCCATTCATGAC 343
QY 241 TGAAGATTACTCATGATAAAGATCCAAACTATAGCTAAAAAGATGAGGATATTTGATGA 300
DB 344 TGAAGATCTCTCATGATAAAGATCCCAATTTATGAGTTGAGAGATTCAGACATTTGTCAATGA 403
QY 301 GGTCAAGGTGATATGTTATCAAGGTAGATGGAATAATCATATGTTTACCTTAAGGATGC 360
DB 404 AATCAAGGTGATATGTTATCAAGGTAGATGGAATAATCATATGTTTACCTTAAGGATGC 463
QY 361 TGGCCACGGGAGTAAGTCCGTACAAAAGAGAAATCATTCGACAAAAACAAGAGATG 420
DB 464 AGTCAAGGTGATATGTTATCAAGGTAGATGGAATAATCATTCGCAAAAAACAAGATG 523
QY 421 TCAACATCTGTAAGGTGAAGTCCAGAAAAAGATGCTGCTGCTGGACGCTTCGCA 480
DB 524 TCAACATCTGTAAGGTGAAGTCCAGCAAAAGATGCTGCTGCTGGACGCTTCGCA 583
QY 481 AAGAGCTTACTACAGATGATGTTATATCTTTAAATGCTTCTGATATCAATGAGATAC 540
DB 584 GGGAGCTTACTACAGATGATGTTATATCTTCAATGATCTGATATCATGAAATGATAC 643
QY 541 TGTGATGCTTATATGCTTCTCATGAGATCATTTACCATTTACATTCCTPAAGATGAT 600
DB 644 GGGAGCTTACTATGCTTCTCATGAGATCATTTACCATTTCTPAAGATGAT 703
QY 601 ATCAGCTAGAGTGGCTGCTGACAGAGCTTCATCTGGTCSGAGAAATCTGCAAA 660
DB 704 ATCAGCTAGAGTGGCTGCTGACAGAGCTTCATCTGGTCSGAGAAATCTGCAAA 763
QY 661 TTCAAGAACCTATCGCCGACAAAATAGCATAACACTTCAGAAACAACTGGGTACTTC 720
DB 764 TTTAAGAACCTATCGCCGACAAAATAGCATAACACTTCAGAAACAACTGGGTACTTC 823
QY 721 TGTAGCAATCCAGAACTACAAATCTAACACAGCAACACAGCAACATPAACGTCA 780
DB 824 TGTAGCAATCCAGAACTACAAATCTAACACAGCAACACAGCAACATPAACGTCA 883
QY 781 AGCAAGTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 840
DB 884 AGCAAGTCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 943
QY 841 TCAACGACATGTAAGATGTAAGTGGCTTGTCTTTGATCCAGCAAAATCAAGTGAAC 900
DB 944 TCAACGACATGTAAGATGTAAGTGGCTTGTCTTTGATCCAGCAAAATCAAGTGAAC 1003
QY 901 AGCTAAGGTGTCAGTGCACAGAGATCATTTACCATTTATCTTACTCTCAAT 960
DB 1004 CGCGAAGGTGTAAGTGCCTCATGTAACCATTTATCTTATGAACAAAT 1063
```

```
QY 961 GTCTGAATTTGGAAGACGATGCTGTAATTAATCCCTTCGTATCGTTCAACCATTTG 1020
DB 1064 GTCTGAATTTGGAAGACGATGCTGTAATTAATCCCTTCGTATCGTTCAACCATTTG 1123
QY 1021 GGTACCAATTTCAAGCCCAAGAAACAAGTCCACACCGACTCCGGAACCTAGTCCAG 1080
DB 1124 GGTACCAATTTCAAGCCCAAGAAACAAGTCCACACCGACTCCGGAACCTAGTCCAG 1183
QY 1081 CCGGACCTGACCAAAATCTTAAATAGACTCAA-----ATTCCTTTTGGT 1128
DB 1184 TCCGCAACCTGACCAAAATCTTAAATAGACTCAA-----ATTCCTTTTGGT 1243
QY 1129 TAGTCAGCTGTACCAAAAGTTGGGGAAGGATATGATTTGGAAGAAAGGCGATCTCTG 1188
DB 1244 CAAGAAGCTGTCCAAAAGTGGGGAAGGATATGATTTGGAAGAAAGGCGATCTCTG 1303
QY 1189 TTAGTCTTTGGGAAGATTTACATCTGAACCTGTAATAAAATCTTGAAGCAAGTTATC 1248
DB 1304 TTAGTCTTTGGGAAGATTTACATCTGTAATAAAATCTTGAAGCAAGTTATC 1363
QY 1249 AAAAGAGAGTGTTCACACACTTAACTGCTAAAAAGAAATGTTGCTCTGCTGA 1308
DB 1364 CAAGCAGAAAGTGTTCACACACTTAACTGCTAAAAAGAAATGTTGCTCTGCTGA 1423
QY 1309 CCAGAAATTTATGATTAAGCATATATCTGTTAAGCTCAATGAAGCTTGTGTTGN 1368
DB 1424 TCGAGAAATTTATGATTAAGCATATATCTGTTAAGCTCAATGAAGCTTGTGTTGN 1483
QY 1369 AAATTAAGGTGTAATTTCTGATTTCCAGCCTTACAGCAAAATTTATGAACGCTTGAATGA 1428
DB 1484 TAATTAAGGTGTAATTTCTGATTTCCAGCCTTACAGCAAAATTTATGAACGCTTGAATGA 1543
QY 1429 TGAATCGACTAATAAGAAAAATGCTGATGATTTATGATGATTTGATGATTTGATGAT 1488
DB 1544 TGTCTCAAGTGTAAAGTCAAGTGTGATGATTTATGCTGCTTTAGTCCGATTCG 1603
QY 1489 CCATCCAGAGCACTTGGCAAAACCAAAATCTCAAAATGATGATTAAGTGAAGCAAGTTCG 1548
DB 1604 TCATCCAGAGCACTTGGCAAAACCAAAATCTCAAAATGATGATTAAGTGAAGCAAGTTCG 1663
QY 1549 TATTCCTCAATTAAGTGAATGATTAACAAAGTCAAGTGTTCATTTTGTATGAACATGA 1608
DB 1664 AGTACCAAGTTCGAGGCAAGTACAAACAGAAACGCTTATCTTGTATGATCGCTGA 1723
QY 1609 TATTAATCGATGTAAGAGATGCAATATGTAACGCTCATATGGCCATATGATCATCTGAT 1668
DB 1724 TATTAATCGATGTAAGAGATGCAATATGTAACGCTCATATGATGATGATGATGATGAT 1783
QY 1669 TGAAGAAAGTACCTTTCTGATTAAGAAAAAGTTCAGCTCAAGCTTATCTTAAGAAAA 1728
DB 1784 TGAAGAAAGTACCTTTCTGATTAAGAAAAAGTTCAGCTCAAGCTTATCTTAAGAAAA 1843
QY 1729 AGTATCTTACCTTCATCTCCAGACAGATGTTTAAGCAAAATCCAACTGAGATGATGC 1788
DB 1844 AGTATCTTACCTTCATCTCCAGACAGATGTTTAAGCAAAATCCAACTGAGATGATGC 1903
QY 1789 AGCAGCTATTTCAATCTGTGTAAGGGGAAAAAGCAATTCATCTGTTTCCATCTTCATA 1848
DB 1904 AGCAGCTATTTCAATCTGTGTAAGGGGAAAAAGCAATTCATCTGTTTCCATCTTCATA 1963
QY 1849 TATGTTGAGCATACAGTTGAGTTAAGAAACGGTAATTTGATTTATCTCATTAAGATCA 1908
DB 1964 TATGTTGAGCATACAGTTGAGTTAAGAAACGGTAATTTGATTTATCTCATTAAGATCA 2023
QY 1909 TTACCATATATTAATTTGCTTGTGTTGATGATCAACATCAAAAGCTCCAAATGCTTA 1968
DB 2024 TTACCATATATTAATTTGCTTGTGTTGATGATCAACATCAAAAGCTCCAAATGCTTA 2083
QY 1969 TACCTTGAAGATTTGTTGCGACATTAAGTACTAGTAAGAACCTGACGAAGCTTC 2028
DB 2084 TACCTTGAAGATTTGTTGCGACATTAAGTACTAGTAAGAACCTGACGAAGCTTC 2143
QY 2029 ACATTTATATGATGATGAGGAGCAATGCCAGTATGATGATGATGATGATGATGATGATGAT 2088
```



```

Db 1338 TGAATGAGATTTTACAAATGAGCTTATGATCTTACGACAGAAATTCACCAAGATTACT 1397
OY 1366 TGAATAATGAGGCTGTAATTTCTGATTTCCAGGCTTACGACAAATTAATTAAGAGCTTGAA 1425
Db 1398 TGAATATAAGGCTGACAAAGTTGATTTTGAAGCTTTGATTAACCTTTGGAAGACTCAA 1457
OY 1426 TGAATGAGTACGATTAATTAAGAAATTTGATGATGATTTTATGGCAATTCCTAGACCAAT 1485
Db 1458 GGAATGCCAAGTGAATGAAGTGAAGTGAATTTCTTGGCTTTGAGCTTCGAT 1517
OY 1486 TACCCATCCAGAGCCATGGCCAAACCAATTTCTCAATTTAGATTAATCTGAAGACGAAT 1545
Db 1518 TCGATATCCAGAACGTTTGAAGAAACCAATGGCCAAATTTACCTACATGATGATGAT 1577
OY 1546 TCGATATTCGATTAATGATGATTAATGATTAATGATTAATGATTAATGATTAATGATTAAT 1605
Db 1578 TGAATGAGCAGAGTTGGCAGGCAAGTACACAGAGAGCGTTTATATCTTTGATCTTG 1637
OY 1606 TGAATATATCAGTATGAGAGAGATGATTAATGATTAATGATTAATGATTAATGATTAATG 1665
Db 1638 TGAATATACAGATGAGAGGAGGATGCTATGTAATCCACATATGACCATAGCCACTG 1697
OY 1666 GATTTGAGAAAGATAGCTTTCTGATTAAGAAAGTTGCAAGCTCAAGCTTAATTAAGA 1725
Db 1698 GATTAATAAGATAGTTTGTCTGAAGCTGAGAGAGCGGACCCAGGCTTAATGCTAAGA 1757
OY 1726 AAAAGGATCTTACCTCCATCTCCAGACGAGATGTTAAGCAATTCACCTGAGATAG 1785
Db 1758 GAAAGGTTGACCCCTCTTCCAGACACATCGATGAGAAATACGAGGCAAAAG 1817
OY 1786 TGCAGAGCTATTTTCAATCTGTTGAAGAGGAAACGAATTCACCTGTTGCACTTC 1845
Db 1818 AACAGAGCTATCTTCAACCGGCTGAAGACAGCTAAGAGGCTCCATGATGATGCTG 1877
OY 1846 AATATGCTTGAAGCATAGTTAGGTTAAACCGGTAATTTGATTAATCTTCAATGAAGA 1905
Db 1878 TTAACATCTTCAATTAATCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1937
OY 1906 TGAATTAACATTAATTAATTTGCTGTTGATGATGATGATGATGATGATGATGATGATGAT 1965
Db 1938 CATTATACCAATCAATCAATTTGAGGTTGACGAGGCTTTTGAAGGCACTGAAGG 1997
OY 1966 CTATACCTTGAAGATTTGTTGCGACGATTAATGATTAATGATTAATGATTAATGATTAAT 2025
Db 1998 GATATCTTGAAGATCTTTTGGCCACTGCAAGTATCTATGTCGAACATCCAAAGCAAG 2057
OY 2026 TCCACATTTTAATGATGATGAGGCAATGCCAGTGAATGATGATGATGATGATGATGATGAT 2085
Db 2058 TCCGATTCAGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2117
OY 2086 CAGTGAAGTCCAAATTAAGAACCTTCAAGCGATG----- 2121
Db 2118 TCAAGCTGATCAATCAACGAAACCAAGCGAGGAGAAACCTCAGACAGAAAC 2177
OY 2122 ----- 2121
Db 2178 TGAAGAAAGACCCCTGAGAGAGAAACCGCAAGCGAGAAACCAAGATCTTCAAAACC 2237
OY 2122 ---AGAGCAGTAGAGAGAAACACCTGCTAGCCAGAGATCCTCAAGTAGAGACTGAA 2178
Db 2238 AACAGAGAACAGAAAGATCAACAGAGAGATCAGAAAGAACCTCAGGTGAGAGCTGAA 2297
OY 2179 AGTAGAGCCCACTCAAGAAAGAGAGATTTTCTGCGAAAGTAGAGGATTTGATGCT 2238
Db 2298 GGTGAGAGAAACCTGAGAGAGCTGAAATTTTCTGGAAGAAATCCAGAGATCCAAATAT 2357
OY 2239 GAAAGCAGTAGCAACAGAACTCTAGCTGTTTACGAATTAATTTGATCTTCAATAT 2298
Db 2358 CAAAGTCAATGCAAGAGAGATCTTCAAGAGATTTAAATTAATTTTATTTGGCAGCCA 2417
OY 2299 GGAATACAAATGATGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2358
Db 2418 GGAACAAATATCTATATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2477

```

```

RESULT 5
US-09-765-272-65
; Sequence 65, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESS: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 65:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2290 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 65:
US-09-765-272-65

Query Match 41.3%; Score 987.6; DB 9; Length 2290;
Best Local Similarity 67.7%; Pred. No. 2.9e-225;
Matches 1481; Conservative 0; Mismatches 645; Indels 60; Gaps 5;

OY 1 TTCTTACAGATTTGGACCTGATCAAGCTTGAAGAGGTTAAGAGAA--TAATCGGTTTC 57
Db 4 TTCTTATACATCTGCTGCTCAACCAAGCTGTCAGGTTAAGAAAGTCTAATCGAGTTTC 63
OY 58 CTATATAGATGAGAAACAGAGGAGCAAGCAAAACGGAATTTGACTCCTGATGAGGTTAG 117
Db 64 TTATATATAGTGTATGATGCTGCTGCTCAAAAGGAGCAAAACCTTGACACAGATGAAGTCA 123
OY 118 CAAGGCTGAAGAAACAAATGCTGAGCAAAATGCTCATCAAGATTAACAGACCAAGCTATGT 177
Db 124 TAAAGAGGAGGAGATCAACGCCGAACAAATGATTAAGATTAAGATTAAGATTAAGATTAAG 183
OY 178 CACTTCAATGAGCAACCACTATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 237
Db 184 GACCTCTGATGAGACCACTTATCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 243
OY 238 CAGTGAAGATTAATCAATGAAGATCCAAACTATTAAGCTTAAGAGATTAATTAATTAATTAAT 297
Db 244 CAGTGAAGATTAATCAATGAAGATCCAAACTATTAAGCTTAAGAGATTAATTAATTAATTAAT 303
OY 298 TGAAGTCAAGAGGATGATTAATTAATCAAGGATGAAGAAATTAATTAATTAATTAATTAATTA 357
Db 304 TGAATCAAGAGGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 363

```

QY 358 TGCTCCCAACGGGATTAACGTCGTAACAAGAGAAATCATTCAGACAAACAGAGACA 417
 Db 364 TGACGCTCATGGATGATTAATTCGCAAAAAGAGATTAAGCTCAGAGACAGAGACG 423
 QY 418 TAGTCAACATCGTGAAGGTGAACCTCAAGAAACATGCTGCTGCTTGGCAGCTTC 477
 Db 424 CAGTCATATATAT-----AACTCAAGACATATATGCTGCTGCGACGACG 474
 QY 478 GCAAGGAGCTATACATAGATGATGTTATCTTTATGCTTGAATCATAGAGA 537
 Db 475 CCAAGGAGCTTATACAGAGATGAGTATATCTTCAATGATCATGATATCATAGAGA 534
 QY 538 TACTGTGATGCTTATATCGTTCTCTCATGAGATCATTACATTCATTCCTAAGATGA 597
 Db 535 CACGGGTATGTTATATCGTTCTCTCAGGGGACCATTAATCATTCATTCCTAAGATGA 594
 QY 598 GTTATCAGCTACGAGTTGGCTGCTGCGAGAGCCTTCTATGCTGCGAGAAATCTGTC 657
 Db 595 GTTATCAGCTACGAGTTAGCTGCTGCGAGAGCCT----- 629
 QY 658 AAATTCAGAAACCTATCGCCGACAAATAGAGATTAACCTTCAAGAACAGCTGGTAC 717
 Db 630 -----ATTGGAATGGAGAGAGAGGATCTCGCTCTTCAAGTTCTAGTTATATGC 681
 QY 718 TTCTGTAGCAATCCAGGAACTCAAAATCTAACAAGCAACAAGCAACACTAAGAG 777
 Db 682 AAATCAGCTCAACCAAGATTGTGAGAACCAACATCTGACTGCTCACTTCA 741
 QY 778 TCAGCAAGTCAAGTATGACATTTGATGTTCTTTGAAACAGCTCTCAAACTGCTTT 837
 Db 742 TCA---AAATCAAGGGGAAACATTTCAAGCCTTTTACGTGATTTGCTAAACCTTT 798
 QY 838 GAGTCAAGCACTATGATGATGAGCTGCTGCTTTGATCCAGACAAATCAGAGTGC 897
 Db 799 ATCAGAACGCAATGAGATCTGATGAGCTTTATTTTCAACCCAGGCAATCAGAGTGC 858
 QY 898 AACAGCTAGAGGTGTTGAGTGCACAGAGATCATTACACTTATCCTTACTCTCA 957
 Db 859 AACCGCAGAGGTGATGCTGCTCTCATGCTGTAACATTAACCTTATCCCTTATGAACA 918
 QY 958 AATGCTGAATTTGGAAGACGAATGCTGCTGTTATTTCCCTTCTGTTATCGTTCAACCA 1017
 Db 919 AATGCTGAATTTGGAAGACGAATGCTGCTGTTATTTCCCTTCTGTTATCGTTCAACCA 978
 QY 1018 TTGGGTACCAATTCAGAGCCGACAAACCAAGTCCACACGATCCGGAACCTAGTCC 1077
 Db 979 TTGGGTACCAATTCAGAGCCGACAAACCAAGTCCACACGATCCGGAACCTAGTCC 1038
 QY 1078 AGGCCCGCACTGCGACCAAAATCTTAAATAGACTCAAAATCTTC-----TTT 1125
 Db 1039 AAGTCCGCACTGCGACCAAAATCTTAAACGCTCCAGCAATCCAAATTTGATGAGAAAT 1098
 QY 1126 GGTATGCTGCTGTAAGAAAGTTGGGAGAGATATGATTTGAGAAAGGCGATCTC 1185
 Db 1099 GGTCAAAAGAGCTTTCAAAAGAGGCGATGATGATGCTTTGAGAGAAATGAGGTTTC 1158
 QY 1186 TCGTTATGCTTTGGGAAAGATTTCCATCTGAAACCTGTTAAATCTTGAAGCAAGTT 1245
 Db 1159 TCGTTATATCCAGCAAGAGATCTTTCAGAGAAACAGAGCAGCATTTGATGAGAAACT 1218
 QY 1246 ATCAAAACAGAGAGTGTTCACACACTTTAACTCTAATAAAGAAATGTTGCTCTCG 1305
 Db 1219 GGCACAGAGAGAAATTTATCTCATAGCTAGAGCTAAGAAATGAGCTCCCATCTAG 1278
 QY 1306 TGACCAAGATTTTATGATTAAGCATATATCTGTTAACTGAGGCTCATAAAGCTTGT 1365
 Db 1279 TGATGAGAAATTTTATCAATAGGCTTATGACTTACTACAGAAATTCACCAAGATTACT 1338
 QY 1366 TGNAAATAGGCTGCTAATTCGATTTCCAGCCTTAACAATATATGAAGCTTGA 1425
 Db 1339 TGATATATAGGCTGACAGTTGATTTTGGGCTTTGATTAACCTGTTGAAACGCTCA 1398
 QY 1426 TGATGATCGACTAATAAAGAAATTTGATGATGATTTATGCAATCTTGAACCAAT 1485

Db 1399 GGATGTGCAAGTATTAAGTCAAGTTAGTGAAGATATCTTCTTCTAGCTCCGAT 1458
 QY 1486 TACCATTCAGAGGACTGTGGCAACCAATTTCTCAATGATATCTGAAGACGAGT 1545
 Db 1459 TCGTATTCAGAACTTTAGAAAAACCAATGCCCCAAATTACTTACACTGATGAGAT 1518
 QY 1446 TCGATTGCTCAATTAAGTATAGTATACAACTCAGATGTTATATTTTGTGAACA 1605
 Db 1519 TCAAGTACCAAGTTGGCAGGCAATACACAGAAAGAGCGTTATATCTTTGATCCCG 1578
 QY 1606 TGATATATCAGTATGAAGAGATGCAATATGTAACCTCATATGAGGCCATGACTG 1665
 Db 1579 TGATATTAACAGTATGAGGGGATGCTTATGTAATCTCAATATGACCCATGACCTG 1638
 QY 1666 GATTGAAAAGATGCTCTTCTGTAAGAAAAAGTTGCGAGCTCAAGCTTATCTAAGA 1725
 Db 1639 GATTAAAAAGATGTTGTTGTTGAGCTGAGAGCGGCGCCAGGCTTATGCTAAGA 1698
 QY 1726 AAAAGTATCTTACCTCATCTCCAGAGCAGATGTTAAAGCAATCCAACTGAGATAG 1785
 Db 1699 GAAAGTTTGAACCCCTCTCTGACAGACATCAGAGATTCAGAAATATCTGAGCAAAAG 1758
 QY 1786 TGACAGCTATTTTACATGCTGTGAAGGGGAAAAAGAAATTCATCTGTTGCACTTC 1845
 Db 1759 AGCAGAACTATCTACAAACGCGTGAAGAGCTAAGAGGTGCACTTGAATCTGATGCC 1818
 QY 1846 ATATATGTTGAGATACAGTACGTTGAGTTAAAAAGGTTATTTGATTTCTCTAAGA 1905
 Db 1819 TTACAAATCTTCAATATATCTGAGAGTCAAAACGTTATATCTCTCATATGA 1878
 QY 1906 TCATTACCATTAATTAATTTGCTGTTGATGATCACACATACAAAGCTCCAAATG 1965
 Db 1879 CCATTACCATTAATTAATTTGAGTGTGTTGAGAGGCTTTATGAGGACCTTAAGGG 1938
 QY 1966 CTATACCTTGAAGATTTGTTGCGAGATTAAGTACTAGTGAACAACCTGACGAG 2025
 Db 1939 GTATACCTTGAAGATCTTTTGGCGAGCTGTCATATCTATGTCMAACATCCAAACGAG 1998
 QY 2026 TCCCATTTCAATATGATGAGGCGCAATGCGCAGTACATGTTGTTAGCAAGAAAGACA 2085
 Db 1999 TCCGATTCAGATATATGTTTGGTAAACGCTAGGACATGTTCAAGAAACAAATATG 2058
 QY 2086 CAGTGAAGATCCAAATTAAGAACTTCAAGCGGATGAAGAGCAGTAGAGAAACACCTGC 2145
 Db 2059 TCAAGCTGATACATCAATCAACGAAACCAAGGAGAGAAACCTCAGACAGAAAAACC 2118
 QY 2146 TGAGCCAGAACTCCCTCAAGTAGAGA 2171
 Db 2119 TGAGGAAGAAACCCCTCGAGAAAGAGA 2144

RESULT 6
 US-09-884-465A-5
 ; Sequence 5, Application US/09884465A
 ; Publication No. US20030077293A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shire Biochem, Inc.
 ; APPLICANT: Hamel, Josee
 ; APPLICANT: Brodeur, Bernard
 ; APPLICANT: Martin, Denis
 ; APPLICANT: Charland, Nathalie
 ; APPLICANT: Ouellet, Catherine
 ; TITLE OF INVENTION: Streptococcus Antigens
 ; FILE REFERENCE: 055190-0044
 ; CURRENT APPLICATION NUMBER: US/09/884,465A
 ; PRIOR APPLICATION NUMBER: 60/212,683
 ; PRIOR FILING DATE: 2000-06-20
 ; NUMBER OF SEQ ID NOS: 384
 ; SOFTWARE: Patent version 3.1
 ; SEQ ID NO 5
 ; LENGTH: 2639

TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-884-465A-5

Query Match 41.1%; Score 980.8; DB 11; Length 2639;
Best Local Similarity 67.8%; Pred. No. 1.3e-223;
Matches 1473; Conservative 0; Mismatches 638; Indels 60; Gaps 5;

1 TTCTTACGAGTTGGAGCTGTATCATCAAGCTAGAAAGGTTAGAAAA---TAATCGTGTTC 57
Db 173 TTCTTACGAGTTGGAGCTGTATCATCAAGCTGTAGAAAAAGCTTAATCGAGTTTC 232
Qy 58 CTATATAGATGAAAAAACAAGCAGCAAGCAAAAGGAAATTTGACTCTGATGAGTTAG 117
Db 233 TTATATAGATGATGATCAGCTGTGTCAAAAGGCAAAAATTTGACCCAGATGAGTCTAG 292
Qy 118 CAAAGCTGAAGAAATTAATGCTGAGCAAAATGCTCATCAAGATTAACAGACCAAGCTATGT 177
Db 293 TAAGAGAGAGGGAGATCAACGCGAACAAATTTGATCAAGATTAACGATCAAGGTTATGT 352
Qy 178 CACTTACATGGGAGACCATATCATTTATTAACAATGTTAGTTCTTATGAGGCTATCAT 237
Db 353 GACCTCTCATGAGACCATTAATCATTAATGAGCAAGTTCTTATGATGCCATCAT 412
Qy 238 CAGTGAAGATTTACTCATGAAAAAGATCAAACTATAAGCTAAAAAGATGAGATATTTTAA 297
Db 413 CAGTGAAGAACTTTCTCATGAAAAAGATCCGAATTAATGATGAGGATTCAGACATTTGTCA 472
Qy 298 TGAAGTCAAGGTTGATATGTTATCAAGTAGATGAAAAATCTATGTTTACCTTAAGA 357
Db 473 TGAATAACAAGGTTGATGATTAAGGTAGACGAAAAATCTATGTTTACCTTAAGA 532
Qy 358 TGCTGCCCCAGCGGATTAACGTCGTATAAGAGGAAATCATCAACAAAAACAAGAGA 417
Db 533 TGCGGCCCATGCGGACAAATTCGACAAAGAAAGATTAACGTCAGAAAGAGAAAGA 592
Qy 418 TAGTCAACATCTGTAAGGTGAATCTCAAGAAACGATGCTGCTGTTGCCCTTGACAGTTC 477
Db 593 CAGTCATATATCAI-----AACTCAAGAGCAATATGCTGTTGCTGACGCCAGAC 643
Qy 478 GCAAGAGCGCTTACTACAGATGATGTTATCTTTAAATGCTTCTGATATCATAGAGA 537
Db 644 CCAAGAGCGTTATACACGATGATGAGGTATCTTCATGATGATCATGATATCAITAGAGA 703
Qy 538 TACTGTGATGCTTATATGTTCTCATGAGAGATCAATCAATTCATTCCTTAAGAATGA 597
Db 704 CACGGGTATGCTTATATGTTCTCATGCGGACCAATTAATCAATTCATTCCTTAAGAATGA 763
Qy 598 GTTATCAGCTAGCAGATTGGCTGCTGACAGAACCTTCTATGCTGTCAGAGAAATCTGTC 657
Db 764 GTTATCAGCTAGCAGATTAGCTGCTGTCAGAAAGCT----- 798
Qy 658 AAATTCAGAACCTTATCGCCGACAAAATAGCAATTAACCTTCAAGAACAACTGGGTACC 717
Db 799 -----ATTGGAATGGAGACAGAGGATCTGCTCCTTCTCAAGTTCTAGTTATTAATGC 850
Qy 718 TTCTGTAAGCAATCAAGAACTAAATTAATCAACACAAACAAACAGAACATCAACAG 777
Db 851 AAATCAGTTCAACCAAGATTTGTCAAGAAACCAATCTGACTGTCTACTCAACTTATCA 910
Qy 778 TCACCAAGTCAAAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 837
Db 911 TCA-----AATCAAGGGGAAAAACATTTCAAGCTTTTACGTGATGATGATGATGATGATG 967
Qy 838 GAGTCAAGCAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTA 897
Db 968 ATCAAGAGCGCATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATGTAATG 1027
Qy 898 AACAGCTGAGGTTGTCAGTGCACAGAGATCAATTCACATTCATTCATTCATTCATTCATTC 957
Db 1028 AACCGCAGAGGTGATGCTGCTCCTCAATGTAACATTAACATTAACATTAACATTAACATTA 1087
Qy 958 AATGTCGTAATGGAAGAAAGAAATGCTGTAATTAATCCCTTCGTTATGTTCAAAACA 1017

1088 AATGTCGTAATGGAAGAAACGAATGCTGTAATTAATCCCTTCGTTATGTTCAAAACA 1147
Qy 1018 TTGGGTACAGATTAAGGCCAGAACCAACAGTCCACACAGCTCCGGAAACCTAGTCC 1077
Db 1148 TTGGGTACAGATTAAGGCCAGAACCAACAGTCCACACAGCTCCGGAAACCTAGTCC 1207
Qy 1078 AGGCCCGAACCTGACCAAAATCTTAATATGACTCAATTCCTC-----TTT 1125
Db 1208 AAGTGTGACCTGACCAAAATCTTAATGACTCAATTCCTC-----TTT 1267
Qy 1126 GGTATGCTGCTGTAGCAAAAAGTTGGGAAAGATATGTAATTCGAAGAAAAGGATCTC 1185
Db 1268 GGTCAAGAAAGCTGTTCAAAAAGTGGGAGGATGATGATGATGATGATGATGATGATGATG 1327
Qy 1186 TCGTATGCTCTTTGGGAAAGATTTACATCTGAAACCTGTAAATCTTTGAAAGCAAT 1245
Db 1328 TCGTATATCCCAAGCAAGATCTTTGACAGAAACACAGCAGGCAATGATGACAACT 1387
Qy 1246 ATCAAAACAGAGAGTGTTCACACACTTAACTCTAAAAAAGAAATGTTGCTCCTCG 1305
Db 1388 GGCACAGAGAGAGATTTATCTCATGACTAGAGCTAAGAAACTGACCTCCATCTAG 1447
Qy 1306 TGACCAAGATTTTATGATTAAGCATATATCTGTTAACTGAGGCTCAATAAGCCTTGT 1365
Db 1448 TGATGAGAAATTTTCAATTAAGGCTTATGACTTATGACAGAAATTCACAAAGATTTACT 1507
Qy 1366 TGNAAATTAAGGTCGTAATCTGATTTCCAGACCTTAAGCAATTAATTAAGACGCTTGA 1425
Db 1508 TGATATTAAGGTCACAAAGTGAATTTGAGTTTGTGATTAACCTGTTGAACGACTCA 1567
Qy 1426 TGATGAATCGACTAATAAAGAAATTTGATGATGATTTATGATGATTTATGATGATGAT 1485
Db 1568 GATGCTCAAGTCAATTAAGTCAATTAAGTCAATTAAGTCAATTAAGTCAATTAAGTCAAT 1627
Qy 1486 TAACCATCAGAGCAGCTTGGCAAAACAATTTCTCAATTTAGTATGATGAGAGCAAGT 1545
Db 1628 TCGTATCAGAAAGCTTTAGGAAACCAATGCGCAATTTACCTACATGATGATGATGAT 1687
Qy 1546 TCGTATGCTCAATTAAGTCAATTAAGTCAATTAAGTCAATTAAGTCAATTAAGTCAAT 1605
Db 1688 TCAAGTACCAAGTGGAGGAGGACATACCAACAAACAGGTTATCTTGTGATCTCG 1747
Qy 1606 TGATATTAACGATGTAAGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1665
Db 1748 TGATATTAACGATGTAAGAGGATGATGATGATGATGATGATGATGATGATGATGATGAT 1807
Qy 1666 GATTGAAAAGATGACCTTTCTGATTAAGAAAGATGCTCAAGCTCAAGCTTACTAAAGA 1725
Db 1808 GATTGAAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1867
Qy 1726 AAAAGTATCTCACTCATCTCCAGAGCAGATTTAAAGCAATTCACATGAGATAG 1785
Db 1868 GAAAGTATGACCCCTCTTGCAGACACACAGATTTAGGAAATTAAGAGGCAAAAGG 1927
Qy 1786 TGACAGACCTTTTCAATCGTGAAGAGGAAAGAAAGAAATTCACATGAGGATGATGATG 1845
Db 1928 AGCAGAGCTATTAACACCGGTGAAGACGCTTAAGAGGTCACATGATGATGATGATGATG 1987
Qy 1846 ATATATGCTGAGCATCAAGTGAAGTTAAACCGTAAATTTGATTAATTCCTCATAAAGA 1905
Db 1988 TTACATCTTCAATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2047
Qy 1906 TCATTAACATTAATTAATTTGCTGTTGATGATCAACATCAAAAGCTCCAAATAG 1965
Db 2048 CCATTAACATTAATTAATTTGAGTGTGATGATGATGATGATGATGATGATGATGATGATG 2107
Qy 1966 CTATACCTTGAAGTTGTTGGAGATTAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 2025
Db 2108 GTATGCTTGAAGTCTTTTGGAGCTGCTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 2167
Qy 2026 TCCACATTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2085

Db 2168 TCCGATTCAGATTAAGTTTGGTAAAGCTAGTACCATGTTCTGTAATAAATAAGGACGA 2227
 Qy 2086 CAGTGAAGATCCAAATAAGAACTTCAAGCGGATGAGAGCAAGCAACCGCC 2145
 Db 2228 CCAGATATGTAACCTGATGAAGATGAGACATGATGAAGTGAAGCCCACTCACCC 2287
 Qy 2146 TGAGCCAGAG 2156
 Db 2288 TGAATCTGATG 2298

 RESULT 7
 US-09-765-272-181
 ; Sequence 181, Application US/09765272
 ; Patent No. US2002061545A1
 ;
 ; GENERAL INFORMATION:
 ; APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESSES:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ;
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ;
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/765,272
 ; FILING DATE: 22-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/961,083
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brookes, A. Anders
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB340P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ;
 ; INFORMATION FOR SEQ ID NO: 181:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1342 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ;
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 181:
 ;
 ; US-09-765-272-181

 Query Match 16.1%; Score 385.4; DB 9; Length 1342;
 Best Local Similarity 65.3%; Pred. No. 1.4e-81;
 Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

Qy 277 AAAGATGAGATATTGTTAATGAGTCAAGGCGATGATGTTATCAAGTAGATGAAA 336
 Db 268 TAAGACCGCTGATATTGTTCAATGAACTCAAGGCTGTTATATCATCAAGTCGATGAAA 327
 Qy 337 ATACTATGTTTACCTTAAGATGCTGCCACCGGATTAACCTCGCTGTAACAAAAGAAAT 396
 Db 328 ATATTATGTTTACCTGAAGATGACGCTCATGCTGATTAATGTTGAAGTAAAGATGAAT 387
 Qy 397 CAATCGCAAAAAACAAGAGCTAGTCAACATCGTGAAGGTGGAACCTCCAGAAACGATG 456
 Db 388 CAATCGTCAAAAAACAAGAGCTAGTCAACATCGTGAAGGTGGAACCTCCAGAAACGATG 435
 Qy 457 TGCTGTGCTTGGCAGCTTGGCAGAGAGCTTACTACTACAGATGATGTTATCTTTAA 516
 Db 436 TAATGTTGCTGTAGCAAGGCTCTGAGGACGATATACGACAAATATGTTATGCTTTAA 495
 Qy 517 TGCTTGTATATCATAGAGATAGTGTGATGCTTATATGCTTCTCATAGAGATCATTA 576
 Db 496 TCCAGCTGATATTTATCGAAGATACGGGTAATGCTTATATGTTCTCATGAGGTCACCTA 555
 Qy 577 CCATTCAATTCCTTAAGATGATGATGATGAGCTAGCGAGTTGGCTGCTGCAAGCTTCT 636
 Db 556 TCACCTACATTCCTCAAAAGCGATTTATCTGCTAGTGAATTAGCAGCAGCTTAAGCAATCT 615
 Qy 637 ATCTGTGAGGAATCTGTCAAATTCAGAACCTATCGCCGACAAATAGCGATTAACAC 696
 Db 616 GCGTGAATAAATATGCAACCGAGTCACTTAATCTTCAACAGCTAGTACAA--- 672
 Qy 697 TTCAAGAACAACTGGGTACCTTCTGTAGCAATCCAGAACTTCAAAATATCAACACAG 756
 Db 673 -----TAACAGCAATCTGT 687
 Qy 757 CAACAAGCAACCTTAACAGTCAAGCAAGTCAAAAGTAAATGATGATGCTCTTGA 816
 Db 688 AGCAAAAGATCAACTGACAGCAGCAAAATTAATCTGAAATCTCAGAGCTTTTGA 747
 Qy 817 ACAGCTTCAAACTGCTTGAAGTCAAGCATGATGAAATCTGATGAGCTTGTCTTGA 876
 Db 748 GGAACCTTATGATTCACCTAGCCGCCCAAGCTTAAAGTAAATGATGAGCTGCTTTGA 807
 Qy 877 TCCAGCAAAATCAACAAGTGAAGTGAAGTGTGTCAGTGCACACGAGATCATTA 936
 Db 808 CCGCTTAAGATTTATGATGCTGATACCAAAATGAGATGCGATTCGATGCGACATTA 867
 Qy 937 CCATTCAATCCCTTACTCTCAATGCTGAATGGAAGAGATGCTGCTATATATCC 996
 Db 868 CCACCTTATTCCTTACAGCAAGCTTCTGCTTGAAGAAAGATGCGAAGATGTGCC 927
 Qy 997 CCTTGT 1003
 Db 928 TATCAGT 934

 RESULT 8
 US-09-769-787-246
 ; Sequence 246, Application US/09769787
 ; Publication No. US20030091577A1
 ;
 ; GENERAL INFORMATION:
 ; APPLICANT: Microbial Technics Limited
 ; APPLICANT: Gilbert, Christophe FG
 ; APPLICANT: Hansbro, Phillip M
 ; TITLE OF INVENTION: Proteins
 ; FILE REFERENCE: PWC/P21129WO
 ; CURRENT APPLICATION NUMBER: US/09/769,787
 ; PRIOR FILING DATE: 2001-01-26
 ; PRIOR APPLICATION NUMBER: GB 9816337.1
 ; PRIOR FILING DATE: 1998-03-27
 ; PRIOR APPLICATION NUMBER: US 60/125164
 ; NUMBER OF SEQ ID NOS: 388
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 246

LENGTH: 1455
 TYPE: DNA
 ORGANISM: Streptococcus pneumoniae
 US-09-769-787-246

Query Match 16.1%; Score 385.4; DB 11; Length 1455;
 Best Local Similarity 65.3%; Pred. No. 1.5e-81;
 Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

```

37 TAAGGAAATATTCGTTTCTTATATGTAAGAAACAGGACCCCAAAACGAGAA 96
99 TAAGGCAATTAATCGTCTCTTATGTGATGCGACCCAGTCAAGTCAAGAAAGTAAAA 158
97 TTTGACTCCTGATGAGTTAGCAAGCGGTAAAGAAATCATGCTGACCAATCGTCATCA 156
159 CTTGACACGACGACGAGTTAGCCAGAAAGAAAGAAATTCAGGCTGAGCAATTTATCA 218
157 GATTAACAGACCAAGGCTATGTCATTCACATGCGACCACTATCATTTATTAATGATA 216
219 AATTACAGATCAGGGCTATGTAACGTCACACGGTGACCACTATCATTTATTAATGATA 278
217 GGTTCCTTATGACGCTATCATGTAAGAAATTAATCAAGAAAGATCCAACTATAAGCT 276
219 AGTTCCTTATGATGCGCTCTTTAGTAAGAACTTTGATGAAGGATCCAACTATCAACT 338
277 AAAAGATGAGATATTTGTAATGAGGTCAAGGCTGATATGTTATCAAGTATGATGAAA 336
339 TAAAGACGCTGATATTTGTAATGAGGTCAAGGCTGATATGTTATCAAGTATGATGAAA 398
337 ATACTATGTTTACCTTAAGAGTCTGCGACGCGATTAACGTCGTAACAAAGAGAAAT 396
399 ATATTATGTTCTACTGTAAGAGTCAAGCTCATGCTGATTAATGTTGTAAGTAAAGTAAAT 458
397 CATATGACAAAACAAAGACATATGTAACATGCTGAAGGTGAACCTCCAAAGAACATG 456
459 CAATGCTCAAAAACAAAGACATGTCAAATATATGAAAGTTA-----ACTC 506
457 TCGTGTGCTTGGACGTTTGGCAAGGCTATTAATGAGATGATGTTATATCTTTAA 516
507 TAAATGTTGCTGAGCAAGGCTTCAAGGAGATATGACAAATGATGTTATGTTCTTTAA 566
517 TCGTCTGATATCATAGAGATATGATGATGTTATGTTATGTTCTCATGAGATCATTA 576
567 TCCAGCTGATATTTATGAAAGATACGGGTATGTTATGTTCTCATGAGGATCATTA 626
577 CCATTCATTTCTTAAGATGATGTTATGATGAGGCTGCTGCGCAAGACCTTCT 636
627 TCACTACATTTCCAAAAGGATTTATCTGATGATTAAGCAGCAGCTTAAAGCATCT 686
637 ATCTGTGAGGAAATCTGCAAAATTCAGAACTATGCGCGACAAATATAGGATACAC 696
687 GGTGGAATAAATATGCAACCGAGTCAGTTAAGCTATTTCTTCAACGCTAGAGCAA--- 743
697 TTCAAGAACAACTGGGTAACCTTCTGTAAGCAATCCAGAACTACAAATATTAACAAG 756
744 -----TAAACGGAATCTGT 758
757 CAACAACAGCAACATTAACAGTCAAGCAAGTCAAAATATGACATTTGATGTTCTTTGA 816
759 ACAAAGATCAACATGCAAGCAGCAAGCAATTAATCTGAAAATCTCCAGACGTTTGA 818
817 ACAGCTCTCAAACTGCTTTGAGTCAAGACATGTAATGATGAGGCTTGTCTTTGA 876
819 GGAATCTATGATTAACCTTAGCGCCCAAGGTTACATGATGATGAGGCTGCTTTTGA 878
877 TCCAGACAAATCAACAGTCAAGCTAGAGGTTGAGTCCACAGAGATCATTA 936
879 CCTGTAGATTAATAGTGTATACCAATGAGAGTTGGATTCCGCAATGGGAGACATTA 938
937 CCATTCATCTCTTACTTCAATGTTCTGATTTGAAAGCAATGCTGCTGATTAATTC 996
939 CCATTTATTTCTTACAGCAAGCTTTGCTTGAAGAAAGATTTGCGAAGATGCTGCT 998
  
```

QY 997 CCTTGT 1003
 DB 999 TATCAGT 1005

RESULT 9
 US-09-769-744A-23
 ; Sequence 23, Application US/09769744A
 ; Publication No. US20030134407A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Le Page, Richard WF
 ; APPLICANT: Wells, Jeremy M
 ; APPLICANT: Hannify, Sean B
 ; APPLICANT: Hansbro, Philip M
 ; TITLE OF INVENTION: Proteins
 ; FILE REFERENCE: PWC/P21122MO
 ; CURRENT APPLICATION NUMBER: US/09/769,744A
 ; PRIORITY FILING DATE: 2001-01-26
 ; PRIOR APPLICATION NUMBER: PCT/G899/02452
 ; PRIOR FILING DATE: 1999-07-27
 ; PRIOR APPLICATION NUMBER: GB 9816336.3
 ; PRIOR FILING DATE: 1998-07-27
 ; PRIOR APPLICATION NUMBER: US 60/125329
 ; PRIOR FILING DATE: 1999-03-19
 ; NUMBER OF SEQ ID NOS: 196
 ; SOFTWARE: PatentIn Ver. 2.1
 ; SEQ ID NO 23
 ; LENGTH: 1455
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 US-09-769-744A-23

Query Match 16.1%; Score 385.4; DB 12; Length 1455;
 Best Local Similarity 65.3%; Pred. No. 1.5e-81;
 Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

```

37 TAAGGAAATATTCGTTTCTTATATGTAAGAAACAGGACCCCAAAACGAGAA 96
99 TAAGGCAATTAATCGTCTCTTATGTGATGCGACCCAGTCAAGTCAAGAAAGTAAAA 158
97 TTTGACTCCTGATGAGTTAGCAAGCGGTAAAGAAATCATGCTGACCAATCGTCATCA 156
159 CTTGACACGACGACGAGTTAGCCAGAAAGAAAGAAATTCAGGCTGAGCAATTTATCA 218
157 GATTAACAGACCAAGGCTATGTCATTCACATGCGACCACTATCATTTATTAATGATA 216
219 AATTACAGATCAGGGCTATGTAACGTCACACGGTGACCACTATCATTTATTAATGATA 278
217 GGTTCCTTATGACGCTATCATGTAAGAAATTAATCAAGAAAGATCCAACTATAAGCT 276
219 AGTTCCTTATGATGCGCTCTTTAGTAAGAACTTTGATGAAGATCCAACTATCAACT 338
277 AAAAGATGAGATATTTGTAATGAGGTCAAGGCTGATATGTTATCAAGTATGATGAAA 336
339 TAAAGACGCTGATATTTGTAATGAGGTCAAGGCTGATATGTTATCAAGTATGATGAAA 398
337 ATACTATGTTTACCTTAAGAGTCTGCGACGCGATTAACGTCGTAACAAAGAGAAAT 396
399 ATATTATGTTCTACTGTAAGAGTCAAGCTCATGCTGATTAATGTTGTAAGTAAAGTAAAT 458
397 CATATGACAAAACAAAGACATATGTAACATGCTGAAGGTGAACCTCCAAAGAAAGATG 456
459 CAATGCTCAAAAACAAAGACATGTCAAATATATGAAAGTTA-----ACTC 506
457 TCGTGTGCTTGGACGTTTGGCAAGGCTATTAATGAGATGATGTTATATCTTTAA 516
507 TAAATGTTGCTGAGCAAGGCTTCAAGGAGATATGCAAAATGATGTTATGTTCTTTAA 566
517 TCGTCTGATATCATAGAGATATGATGATGTTATGTTATGTTCTCATGAGATCATTA 576
567 TCCAGCTGATATTAATGAGATGAGGCTTATGATGATGTTATGTTCTCATGAGGATCATTA 626
577 CCATTCATTTCTTACAGCAAGCTTTGCTTGAAGAAAGATTTGCGAAGATGCTGCT 636
  
```

```

Db      627 TCACACATTCGCCAAAGCGATTTATCTGTAGTGAATTAGACGAGCTAAAGCACATCT 686
Qy      637 ATCTGTGAGGAATCTGTCAAAATTCAGAACCTATGCCGACAAATATGCGATTAAC 696
Db      687 GGCTGGAAAAATATGCAACCGAGTCAGTTAACTATTCTTCACAGCTAGTGACAA--- 743
Qy      697 TTCAAGAACAAACTGGGTACCTTCTGTAGCAATCCAGAACTACAATACTAACACAAAG 756
Db      744 -----TAAACGCAATCTGT 758
Qy      757 CAACACAGCAACTACTAAAGTCAAGCAAGTCAAAAGTAATGACATTGATGCTCTTGAA 816
Db      759 AGCAAAAGATCACTACGACCAAGCCAGCAAAATTAATCTGAAATCTCCAGAGCTTTTGA 818
Qy      817 ACAGCTACAAACCTGCTTTGAGTCAAGCATGTAAATCTGATGCGCTTGTCTTGA 876
Db      819 GGAATCTATGATTCACCTAGCGCCCAAGTTACAGTAATCAGATGAGCTGTGCTTTGA 878
Qy      877 TCCAGCACAATCAACAAGTGCAGACGTAAGAGTGTGACAGCCAGCAGAGATCATTA 936
Db      879 CCCTGTAGATTAATCAAGTGCACCAATGAGATTGCGATTCGCGATGGCGACATTA 938
Qy      937 CCACCTTATCCCTTACCTCAAAATGTCTGAATGGAAGACGAATGCTGTATTATTC 996
Db      939 CCACCTTATTCCTTACGACCAAGCTTTCTGCTTAGAAGAAAGATTGCCAGATGTGCC 998
Qy      997 CCTTCGT 1003
Db      999 TATCAGT 1005

```

RESULT 10

```

US-09-884-465A-1
; Sequence 1, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 3120
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-1

```

```

Query Match 16.1%; Score 385.4; DB 11; Length 3120;
Best Local Similarity 65.3%; Pred. No. 2.3e-81;
Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

```

```

Qy      37 TAAAGAAATAATATGCTGTTCTTATATGATGAAAGAAAGAGGAGCAAGGAGAA 96
Db      99 TAAAGACAATAATATGCTGTTCTTATGATGAGGAGCAAGTCAAGGAGGAGAA 158
Qy      97 TTGATCTCTGATGAGTTAGCAAGCGTGAAGGAATCATGTGAGCAATGCTATCAA 156
Db      159 CTGACACACAGACAGGTTAGCGAGAAGAAAGAAATTCAGGCTAGCAAAATGTAATCA 218
Qy      157 GATAACAGACCAAGGCTATGTCATCTTCAATGCGACACATATCTTTTCAATGTGTA 216
Db      219 AATTACAGATCAGGGCTATGTAAGCTCACAGCGTGACCACTATCTATTAATGGGAA 278

```

```

Qy      217 GGTTCCTTATGACCGTATCATCACTGAAGAATTACTGATGAAGATCCAACTATAAGCT 276
Db      279 AGTTCTTATGATGATCCCTCTTTAGTGAAGAACTCTTATGAGGATCCAACTATCACT 338
Qy      277 AAAAGATGAGATATTTGTTAATGAGTCAAGGCTGATATGTTATCAAGTAAATGGA 336
Db      339 TAAAGACGCTGATTTTGTCAATGAAGTCAAGGGGTGTTAATATCAAGGTCGATGAA 398
Qy      337 ATACTATGTTTACCTTAAAGATGCTGCCACGCGGATTAAGTCGTCGTCGTCGTCGTCG 396
Db      399 ATATTATGTTTACCTTGAAGATGCTGACGTCATGCTGATTAATGTTGAACTAAAGTGA 458
Qy      397 CAATCGCAAAAACAGAGCATAGTCAACATGTAAGTGAAGTCACTCCAGAAAGATG 456
Db      459 CAATGTCAAAACAGAAACATGTCAAAGATTAAGAGGTTA-----ACTC 506
Qy      457 TGCTTTGCTTGGCAGCTTGCAGAGACGCTATATCAATGATATGTTATATCTTAA 516
Db      507 TAAATGCTGTACCAAGGCTCTCAGGAGCATATACAGCAAAATGATGTTATGCTTTAA 566
Qy      517 TGCTTCTGATATCATAGAGTACTGGTATGCTTATATGCTTCTCATGAGATCATTA 576
Db      567 TCCAGCTGATATATTCGAAGTATCGGTAATGCTTATATGCTTCTCATGAGAGCTACTA 626
Qy      577 CCATTACATTCCTAAGATGATATCAGCTAGCAGTGGCTGTCAGAAAGCTTCT 636
Db      627 TCACATCATTCGCCAAAGCGATTTATCTGTAAGATTAAGCAGCAGCTAAAGACATCT 686
Qy      637 ATCTGTGAGGAATCTGTCAATTTCAAGACCTATCGCCGACAAATAGCGTAACAC 696
Db      687 GGCTGGAATAATATGCAACCGAGTCAGTTAAGCTATCTTCAACACCTAGTGCA--- 743
Qy      697 TTCAAGAACAACTGGGTACCTTCTGAAGCAATCCAGAACTCAATATCAACACAG 756
Db      744 -----TAAACGCAATCTGT 758
Qy      757 CAACACAGCAACTACTAAGTCAAGCAAGTCAAAAGTAATGACATTGATGCTCTTGA 816
Db      759 AGCAAAAGATCAACTAGCAAGCCAGCAAAATTAATGAAATCTCAGAGCTTTTGA 818
Qy      817 ACAGCTTCAAACTGCTTTGAGTCAAGCATGTGAATCTGATGCGCTTGTCTTGA 876
Db      819 GGAATCTATGATTCACCTAGCGCCCAAGCTTACAGTAATCAAGATGCGCTGTGTA 878
Qy      877 TCCAGCAAAATCAACAAGTGCAGACGTAAGAGTGTGACAGCCAGCAGAGATCATTA 936
Db      879 CCCTGTAGATTAATCAAGTGCACCAATGAGATTGCGATTCGCGATGGCGACATTA 938
Qy      937 CCACCTTATCCCTTACCTCAAAATGTCTGAATGGAAGAAAGATCGCTGATATTC 996
Db      939 CCACCTTATTCCTTACGACCAAGCTTTCTGCTTAGAAGAAAGATTGCCAGATGTGCC 998
Qy      997 CCTTCGT 1003
Db      999 TATCAGT 1005

```

RESULT 11

```

US-09-884-465A-2
; Sequence 2, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683

```

; PRIOR FILING DATE: 2000-06-20
 ; NUMBER OF SEQ ID NOS: 384
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 2
 ; LENGTH: 5048
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 US-09-884-465A-2

Query Match 16.1%; Score 385.4; DB 11; Length 5048;
 Best Local Similarity 65.3%; Pred. No. 2.9e-81;
 Matches 631; Conservative 0; Mismatches 276; Indels 60; Gaps 2;

```

QY 37 TAAGAAAATATCGTGTCTTATATAGTGAAGAAACGACGCAAAACGAGAA 96
DB 1875 TAAGACAAATATCGTGTCTTATATAGTGAAGTGCACCGCTCAAGTCAGAAAGTGA 1934
QY 97 TTTGACTCTGATGAGTTAGCAAGCGTGAAGAAATCAATGCTGACCAATCTCATCA 156
DB 1935 CTGACACCAAGACGAGTTAGCCGAAAGAAATTCAGGCTGAGCAAAATTTATCA 1994
QY 157 GATTAACAGCAAGGCTATGCTCACTTCACTGCGACCCCTATCATTTATTAAGTGA 216
DB 1995 AATTACAGATCAGGCTATGTAAGTCAACGCTGACCACTATCATTAATATGGA 2054
QY 217 GGTTCCTTATGACGCTATCATCAGTGAAGATTACTCATGAAGATCCAAACTATAAGCT 276
DB 2055 AGTTCCTTATGATGCCCTCTTTAGTGAAGACTCTTGATGAAGATCCAAACTATCACT 2114
QY 277 AAAAGATGAGATATTGTTAATAGATCAAGGCTGATATGTTATCAAGGTAGATGAAA 336
DB 2115 TAAAGCGCTGATATGTCATATGAAGTCAAGGCTGTTATATCATCAAGGTGATGAAA 2174
QY 337 ATACTATGTTTACCTTAAGATGCTGCCACGCGGATACGTCCTGACAAAGAAAT 396
DB 2175 ATATTATGCTACTTAAGATGCTCACTGCTGATTAATGTTGAACTTAAGATGAAAT 2234
QY 397 CAATGACAAAAACAAGAGATGTCACATCGTGAAGTGAATCCAAAGAACATGAG 456
DB 2235 CAATGCTCAAAAACAAGACATGTCAAAGATATGAAAGTTA-----ACTC 2282
QY 457 TCGTGTCTTGGCAGCGTTGCGAAGACGCTATACTACAGATGATGTTATCTTTAA 516
DB 2283 TAAATGCTGTAAGCAAGGCTCAGGAGACGATATACAAATGATGTTATCTTTAA 2342
QY 517 TCGTTTGTATATAGAGATATCTGATGCTTATATGCTTCTCATGAGATCATTA 576
DB 2343 TCCAGCTGATATTATGAAGATACGGGTATGCTTATATGCTTCCATGAGAGTCACTA 2402
QY 577 CCATTACATTCCTTAAGATGATGATAGCTAGCGAGTGGCTGCGAAGCCTTCCT 636
DB 2403 TCACTACATTCCTCAAAAAGCATTTATCTGCTAGTGAATTAAGCAGCTTAAGACATCT 2462
QY 637 ATCTGTGAGAGAAATCTGTCAAAATCAAGAACCTATGCGCGACAAATATAGCATTAAC 696
DB 2463 GCGTGAATAAATATGCAACCGAGTCAGTTAAGCTATTTCTTAAGCGCTAGTACAA---- 2519
QY 697 TTCAAGAACAAACTGGGTACTTCTGTAGCAATCCAGAACCTACAAATCTAACCAAG 756
DB 2520 -----TAAACCGCAATCTGT 2534
QY 757 CAACAACAGCAACTATACAGTCAAGCAAGTCAAGAAATGACATGATGATCTTTGA 816
DB 2535 AGCAAAAGATCAACTAGCAAGCAGCAAAATTAATCTGAATCTCCAGAGTCTTTTGA 2594
QY 817 ACAGCTCTACAACTGCTTTAGTCAACGACATGTAGATCTGATGCGCTGTCTTTGA 876
DB 2595 GGAAGCTCTATGATTAAGCTAGCGCCCAAGTTACAGTGAATGAGTGGCTGTCTTTGA 2654
QY 877 TCCAGACAAATACAAAGTGAACAGCTAGAGGTGTTGACAGTCCACAGGAGATCATTA 936
DB 2655 CCTGTGAATATATCATGATGTCACACCAATGAGATTGCGCATGCGACATTA 2714
  
```

```

QY 937 CCACCTCATCCCTTACTCTCAAAATGCTGAATTTGAAGCAAGATGCTGATTAATTC 996
DB 2715 CCACCTTATCTCTTACAGCAAGCTTCTGCTTTAGAGAAAGATGCGCAAGATGCTGC 2774
QY 997 CCTTGT 1003
DB 2775 TATCACT 2781
  
```

RESULT 12

US-09-884-465A-9
 ; Sequence 9; Application US/09884465A
 ; Publication No. US20030077293A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shire Biochem, Inc.
 ; APPLICANT: Hamel, Josee
 ; APPLICANT: Brodeur, Bernard
 ; APPLICANT: Martin, Denis
 ; APPLICANT: Charland, Nathalie
 ; APPLICANT: Ouellet, Catherine
 ; TITLE OF INVENTION: Streptococcus Antigens
 ; FILE REFERENCE: 055190-0044
 ; CURRENT APPLICATION NUMBER: US/09/884,465A
 ; PRIOR FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: 60/212,683
 ; PRIOR FILING DATE: 2000-06-20
 ; NUMBER OF SEQ ID NOS: 384
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 9
 ; LENGTH: 2528
 ; TYPE: DNA
 ; ORGANISM: Streptococcus pneumoniae
 US-09-884-465A-9

Query Match 14.5%; Score 347; DB 11; Length 2528;
 Best Local Similarity 73.0%; Pred. No. 3e-72;
 Matches 465; Conservative 0; Mismatches 160; Indels 12; Gaps 1;

```

QY 37 TAAGAAAATATCGTGTCTTATATAGTGAAGAAACGACGCAAAACGAGAA 96
DB 39 TAAGACAAATATCGTGTCTTATATAGTGAAGTGCACCGCTCAAGTCAGAAAGTGA 98
QY 97 TTTGACTCTGATGAGTTAGCAAGCGTGAAGAAATCAATGCTGACCAATCTCATCA 156
DB 99 CTGACACCAAGACGAGTTAGCCGAAAGAAATTCAGGCTGAGCAAAATTTATCA 158
QY 157 GATTAACAGCAAGGCTATGCTCACTTCAATGCGACCACTATCATTTATTAAGTGA 216
DB 159 AATTACAGATCAGGCTATGTAAGTCAACGCTGATCATTAATTAATGGA 218
QY 217 GGTTCCTTATGACGCTATATCATGGAAGATTTACTCATGAAGATCCAAACTATPAGCT 276
DB 219 AGTTCCTTATGATGCCCTCTTTAGTGAAGAACTCTTGAAGAGATCCAAACTATCACT 278
QY 277 AAAAGATGAGATATTGTTAATAGGTCGAAGGTCGATGTTATCAAGTGAAGTGA 336
DB 279 TAAAGACGCTGATATTGTTAATGAAGTCAAGGTCGTTATATCATCAAGTGAAGTGA 338
QY 337 ATACTATGTTTACCTTAAGATGCTGCCACGCGATTAACGTCGTACAAAAAGAGAAAT 396
DB 339 ATATTATGCTACTGAAAGATGACGCTCATCTGATTAATGTTGAACTTAAGATGAAT 398
QY 397 CAATGACAAAAACAAGACATAGTCAACATGCTGAAGTGAATCTCCAAAGAACATG 456
DB 399 CAATGCTCAAAAACAAGAACATGTCAAAGATATAGAGGTTAATCTTAA----- 449
QY 457 TCGTGTGCTTGGCAGCGTTGCGAAGACGCTATATCAAGATGATGTTATATCTTTAA 516
DB 459 ---TGTGCTGTAAGCAAGTCTCAGGACGATTAACGACAAATGATGTTATGTTTAA 506
QY 517 TCGTCTGATATCATAGAGATATGCTGATGCTTATATGTTCTCATGAGATCATTA 576
DB 507 TCCAGCTGATATTATGCAAGATACGGGTATATGCTTATATGCTTCCATGAGAGTCACTA 566
  
```


Db 486 AACTAAGAGCTAAAGAAAAGTTTACCTCAAGTGGCCCATCTCAGTAAAGAAAGAT 545
QY 454 TGGTCTGTTGGCTTGGACAGCTTCCGACAGCTTACTACAGATGATGTTATATCTT 513
Db 546 TCGGCGAGTCAATGAAGCAAAAAGACAAAGAGCGTACTACTACAGAGATGGCTATATTTT 605
QY 514 TAATGCTTCTGATATCATAGAGATATCTGATGCTTATATGTTCTCATGAGATCA 573
Db 606 TAGTCCGACAGATATCATATGATGATTTAGAGATGCTTATAGTACCTCATGTAATCA 665
QY 574 TTACCATTTACATTTCTTACAGATGATTTACAGTACGAGTGGCTGCTGCAAGACCTT 633
Db 666 CTATCATTTATATCTCTAAAAAAGATTTCTCAGAGTGAAGTGAAGTGTGCTGCAAGACCTA 725
QY 634 C 634
Db 726 C 726

RESULT 15

US-09-252-088-13
Sequence 13, Application US/09252088
Publication No. US20030031682A1
GENERAL INFORMATION:
APPLICANT: BRODEUR, Bernard R.
APPLICANT: RIOUX, Clement
APPLICANT: BOYER, Martine
APPLICANT: CHARLEBOIS, Isabelle
APPLICANT: HAMEL, Jose
APPLICANT: MARTIN, Denis
TITLE OF INVENTION: NOVEL GROUP B STREPTOCOCCUS ANTIGENS
FILE REFERENCE: 8331-9002
CURRENT APPLICATION NUMBER: US/09/252, 088
CURRENT FILING DATE: 1999-02-18
EARLIER APPLICATION NUMBER: US/60/075,425
EARLIER FILING DATE: 1998-02-20
NUMBER OF SEQ ID NOS: 44
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 13
LENGTH: 5215
TYPE: DNA
ORGANISM: group B streptococcus
FEATURE:
NAME/KEY: CDS
LOCATION: (3)..(122)
FEATURE:
NAME/KEY: CDS
LOCATION: (133)..(2511)
FEATURE:
NAME/KEY: CDS
LOCATION: (367)..(2511)
FEATURE:
NAME/KEY: CDS
LOCATION: (12716)..(2946)
FEATURE:
NAME/KEY: CDS
LOCATION: (2995)..(3252)
FEATURE:
NAME/KEY: CDS
LOCATION: (3299)..(3676)
FEATURE:
NAME/KEY: CDS
LOCATION: (3837)..(4124)
FEATURE:
NAME/KEY: CDS
LOCATION: Complement((4351)..(5214))
US-09-252-088-13

Query Match 8.8%; Score 209.8; DB 11; Length 5215;
Best Local Similarity 60.4%; Pred. No. 2,6e-39;
Matches 399; Conservative 0; Mismatches 232; Indels 30; Gaps 2;

QY 4 TTACGAGTTGGAGTGTATCAAGTAGAAGCGTTTAAAGAAAATATCGTGTCTTATAT 63
Db 111 TTACCAACTGGTGTAGATCATATATGCGTCTAGCAACAAAGGACATCATGATGCTATAT 170
QY 64 AGATGAAAAAAGGACGCAAAAAA-----ACGGAATTGACTCCTGATGAGGTTAG 117
Db 171 TGATGACGCAAAAGTTAAGGCAAAAGCCCTTAAACAAACAAAGATGATCAATCAG 230
QY 118 CAAGGCTAAGGAATCAATGCTGAGCAAAATGCTATCAAGATTAAGACCAAGGCTATGT 177
Db 231 TCGTGAAGAGGCACTCTGCTGAAACAGATGATCTCAAAATATCTGACCAAGGCTATGT 290
QY 178 CACTTCACATGCGACCACTATCATTTATTAACATGTAAGGTTCTTATGACGCTATCAT 237
Db 291 GACCTCAACAGGTGACATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 350
QY 238 CAGTGAAGATTTACTCATGAAAGATCCAAACTATTAAGCTAAAGATGAAGATATTGTTAA 297
Db 351 TAGTGAAAGATTGTTGATGACGAGATCCTAATTAACGTTTAAACATCAGAGTTATCA 410
QY 298 TGAAGTCAAGGATGATATGTTATCAAGTAGATGAGAAATATCTATGTTTACCTTAAGGA 357
Db 411 TGAATCTTTAGACGTTACGTTATTAAGTCAATGGAACATTAATGTTTACCTCAAGCC 470
QY 358 TGTGCCCCCAGCGGATAAGTCCGTACAAAGAGAAATCAATCGACAAACAAAGAGCA 417
Db 471 AGGTGTAAGGCGCAAAACATTTCCAAACCAACAAATTTCTGAGCAAGTAGCAAGG 530
QY 418 TAGTCAACATCGTGAAG-----GTGGAATCCCAAGAAAGCA 453
Db 531 AACTAAAGAACTAAAGAAAAGGTTAGTCAAGTGGCCCATCTCACTAAAGAAAGAT 590
QY 454 TGGTCTGTTGGCTTGGACGTTCCGACGAGAGCGTATCTACAGATGATGTTATATCTT 513
Db 591 TCGGCGACTCAATGAAGCAAAAAGCAAGAGCGCTATACAGAGATGCTATATATTTT 650
QY 514 TAATGCTTCTGATATCATAGAGATATCTGAGATGCTATATGCTTCTCATGAGATCA 573
Db 651 TAGTCCGACAGATATCATATGATGATTTAGAGATGCTTATTTAGTACCTCATGTAATCA 710
QY 574 TTACCATTTACATTTCTTAAAGATGAGTTATCAGCTAGCAGTTGCTGTCAGAAAGCTT 633
Db 711 CTATCATTTATATCTCTTAAAGAAAGATTGCTTCCAAAGTAGAGTGTGTCGCAAGACCTA 770
QY 634 C 634
Db 771 C 771

RESULT 16

US-09-884-465A-257
Sequence 257, Application US/09884465A
Publication No. US20030077293A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 257
LENGTH: 819
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:

OTHER INFORMATION: Unknown Organism
US-09-384-465A-257

Query Match 8.4%; Score 201.2; DB 11; Length 819;
Best Local Similarity 55.8%; Pred. No. 1.1e-37;
Matches 414; Conservative 0; Mismatches 313; Indels 15; Gaps 1;

```

1645 TCATATGGCCCATGCTACGATGTAAGAAAAGATAGCCCTTCTGATTAAGAAAAGTTGC 1704
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 72 TTAATCTTTGATAGTACTGATTTAAAAAGATGTTGTTCTGAGAGCGAGAGCGGC 131
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1705 AGCTCAACCTTACTTAAGAAAAGATGTTCTTCCATCTCCAGACGAGATGTTAA 1764
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 132 AGCCAGGCTTATGCTAAAGAAAAGATTTGACCCCTCTTCAGACAGACCAAGATTC 191
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1765 AGCAATATCCATGAGATAGTACGAGCTATTACAAATGCTGGAAGGGAAGAAAACG 1824
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 192 AGAAATATCTAGAGAAAAGAGAGAGATTAACAACCGCGTGAAGAGAGCTAAAGAA 251
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1825 AATTCACCTGCTTCCATATATGATGAGCATACAGTGAAGTTAAACGGTAA 1884
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 252 GGTGCACTTGTATGCTATGCTTACATCTTCAGATCTGTAGAAAGTCAAAAAGGTAG 311
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1885 TTGATTAATCTCTCAATAGATCATTAATTAATTAATTTGCTGTTGATGATCA 1944
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 312 TTTAATCATACCTCATTTATGACATTACATTAACATCAAAATTTGAGTGTGAAGAG 371
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1945 CACATACAAAGCTCCAAATGCTATACCTTGAAGATTTGTTGAGAGATTAAGTACTA 2004
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 372 CCTTATAGGACACTTAAGGGGTATAGTCTGAGAGATCTTTTGCGACTGTCAAGTACTA 431
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 2005 CGTAGAACACCTCGACAGACGTCACATTTCTAATGATGATGGGGCAATGCGAGTAGCA 2064
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 432 TGTGAAACCGGGAACGCTAATGACCATCTTCTGTAATAAGAGACCAAGATAGTAA 491
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 2065 TGTGTTAGGCAAGAAAGCAACAGTGAAG-----ATCCAAATTAAGAACTT 2109
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 492 ACTGATGAGATGAAGAAACATGATGAAGTAAAGTGAAGCAACTCAGCTGATCTGATGA 551
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 2110 CAAGCGGATGAAGAGCCAGATGAGAGAAAACACTGCTGAGCCAGAAATCCTCAAGTGA 2169
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 552 AAAAGAGATCAACGCTGTTTAAATCTTCAGCAATATATCTTTTAAACCAAGACACTGA 611
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 2170 GACTGAAAAAGTAGAAGCCCAACTCAAGAAAGCAGAAATTTGCTTGGAAAAGTAAACGA 2229
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 612 TACGAAAGAGACAGAGAAAGCTGAAGATATCCACATGAGGTGAATTTCCGTGATC 671
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 2230 TTCTAGTGAAGCAATGCAACAGAAACTCTAGCTGGTTTACGAAATTAATTTGACTCT 2289
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 672 CCTAGTATTTAGACAAATGCTATGAGACATTTGACTGTCTTAAAGATAGCTCTTCT 731
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 2290 TCAAAATTTGATATCATATGATATGAGAGAGCAAGAAAATTAATTTGCTTGTAA 2349
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 732 CGGAACGAAAGATTAACATATTTTCAAGCAGAGATAGTCTCTTGCGCTTTGTAA 791
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 2350 AGAAGTATCTCTTATCTGTA 2371
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 792 AGAAAGTCAACCGCTCTATA 813
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 17

US-09-769-736-71
Sequence 71, Application US/09769736
Publication No. US20030138775A1
GENERAL INFORMATION:
APPLICANT: Microbial Technics Limited
APPLICANT: Le Page, Richard WF
APPLICANT: Wells, Jeremy M
APPLICANT: Hanniffy, Sean B
TITLE OF INVENTION: Proteins
FILE REFERENCE: PWC/P21085WO
CURRENT APPLICATION NUMBER: US/09/769,736
CURRENT FILING DATE: 2003-02-14

PRIOR APPLICATION NUMBER: GB 9816335.5
PRIOR FILING DATE: 1998-07-27
PRIOR APPLICATION NUMBER: US 60/125163
PRIOR FILING DATE: 1999-03-19
NUMBER OF SEQ ID NOS: 212
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 71
LENGTH: 1455
TYPE: DNA
ORGANISM: Streptococcus agalactiae
US-09-769-736-71

Query Match 4.9%; Score 117.4; DB 12; Length 1455;
Best Local Similarity 53.8%; Pred. No. 1.5e-17;
Matches 306; Conservative 0; Mismatches 226; Indels 27; Gaps 2;

```

1469 GCATTCTAGACCAATTACCATTCAGAGGACTTGGCAAAACCAATTTCTCAATTGAG 1528
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229 GAAACCATGACAGTAAACCAAGAAAGACCAACGTAAAGGCAAAACCAATAGCCAGATTGTC 288
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1529 TATATGAGACGAGTTGCTATTGCTCAATTTAGCTGATTAAGTATACAGCTGAGATGCT 1588
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 289 TACAGTCTCAAGAAATTTGAAGGCAAAAAGCTGTTAATACACAACATCTGATGCT 348
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1589 TACATTTTATGATGAACATGATATATATCATGATGAAGAGATGATGTAACGCTCAT 1648
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 349 TACATTTTATGATGATTAAGATATTAATAAAGATACAGGTACAGGTATGATCATTCACAT 408
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1649 ATGGCCATATGCTACTGAGATTGCAAAAAGATAGCTTTCTGATTAAGAAAAGTTGCGCT 1708
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 ATGACATATGACATTTGGGTACCAAAAGAAATTTATCAAGTCTGAAATTAAGACAGCT 468
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1709 CAAGCTATATCTAAGAAAAGATGATCTACCTCATCTCAGACGAGATGTTAAAGCA 1768
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 469 CAAGATTTCTTTAGAAA-----TGTGAAGCAATCAAGACAAA 510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1769 AATCAACTGAGATAGTACGACGACTATTTTACATGCTGTGAAGGGGAAAACGAATT 1828
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 511 CCAAAAACAGTAAACAGCTCAAGAAATCTATGAGCAATTAAGCAAAAGCAATTGTT 570
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1829 CCACTGTTGACTTCATATATGTTGAGATACAGTGAAGTTAAACCGTAAATTTG 1888
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 571 AAACCTGAAGATTTATTTTGTGAATTGCAACGCAAGCACTAATGAATGTATCAATTT 630
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1889 ATTAATCTCATAGATCATTAACATATATTAATTTGCTGTTGATGATCACACA 1948
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 631 GTAAATCTCATTAAGATCATTAACATTAATGTGAATTAATTAATGTTGATGAAGAAA 690
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 1949 TACA-----AAGCTCCAAATGCTATACCTTGAAGATTTGTTGCGACGATTAAG 1999
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 691 GATCTTTAGCTGATTCAGATTAAGACATATTTCTTTAAGACATTTTAGCTACGGCTAAA 750
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Oy 2000 TACTAGTAGAACACCTGACGAACGTC 2028
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 751 TATTACATGATGACCCAGAAAACGTC 779
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 18

US-09-452-599-34/C
Sequence 34, Application US/09452599
Patent No. US20020055101A1
GENERAL INFORMATION:
APPLICANT: Bergeron, Michel G.
APPLICANT: Ouellette, Marc
TITLE OF INVENTION: Specific and Universal Probes and Amplification Primers
TITLE OF INVENTION: to Rapidly Detect and Identify Common Bacterial
TITLE OF INVENTION: Pathogens and Antibiotic Resistance Genes from Clinical
TITLE OF INVENTION: Specimens for Routine Diagnosis in Micro
FILE REFERENCE: 12287.31
CURRENT APPLICATION NUMBER: US/09/452,599
CURRENT FILING DATE: 1999-12-01
PRIOR APPLICATION NUMBER: 08/526,840

PRIOR FILING DATE: 1995-09-11
PRIOR APPLICATION NUMBER: 08/304,732
PRIOR FILING DATE: 1994-09-12
NUMBER OF SEQ ID NOS: 177
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 34
LENGTH: 841
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-09-452-599-34

Query Match 4.6%; Score 109; DB 9; Length 841;
Best Local Similarity 65.4%; Pred. No. 1.1e-15;
Matches 176; Conservative 0; Mismatches 90; Indels 3; Gaps 1;

QY 2090 GAAGATCCAAATAGACTTCAAGCGGATGAAAGCCGATGAGAGAAACCTGCTGAG 2149
DB 633 GAGTCTCCAAACCAACAGAGAGACCGAGAAAGAAATCCAGAAATCCACAGAGAA 574
QY 2150 CCAGAGTCCCTCAAGTAGAGACTGAAAGTAGAAGCCCACTCAAGAGCAGAGATT 2209
DB 573 TCAGAGAAACCTCAGGTGAGACTGAAAGTTAAGAA---AACTGAGAGGCTGAAGAT 517
QY 2210 TTGCTTGCCAAAGTAAAGGATTTGCTGAAAGCCCAATGCAACAGAACTTACGTGT 2269
DB 516 TTAAGTGGAAATCCAGATCCCAATTTATCAAGTCCCAAGAGAGACTCTACAGAGA 457
QY 2270 TTACGAATTAATTTGACTCTTCAATTTAGTAAATAGTATCAATGTCAGAGAGAGA 2329
DB 456 TTTAAATAATTAATTTCTATTGTCACCCAGAGCAAAATATATATGACAGAGCTGAA 397
QY 2330 AAATTACTGCGTGTGTTAAAGGAAGTAA 2358
DB 396 AAATATTGCGCTTTATTAAGAGAGTAA 368

RESULT 19

US-10-121-120-34/C
Sequence 34, Application US/10121120
Publication No. US20030180733A1
GENERAL INFORMATION:
APPLICANT: Bergeron, Michel G.
APPLICANT: Ouellette, Marc
APPLICANT: Roy, Paul H.
TITLE OF INVENTION: Specific and Universal Probes and Amplification
TITLE OF INVENTION: to Rapidly Detect and Identify Common Bacterial
TITLE OF INVENTION: Pathogens and Antibiotic Resistance Genes from Clinical
TITLE OF INVENTION: Specimens for Routine Diagnosis in Micro
FILE REFERENCE: 12287.31
CURRENT APPLICATION NUMBER: US/10/121,120
CURRENT FILING DATE: 2002-04-11
PRIOR APPLICATION NUMBER: 09/452,599
PRIOR FILING DATE: 1999-12-01
PRIOR APPLICATION NUMBER: 08/304,732
PRIOR FILING DATE: 1994-09-12
NUMBER OF SEQ ID NOS: 177
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 34
LENGTH: 841
TYPE: DNA
ORGANISM: Streptococcus pneumoniae
US-10-121-120-34

Query Match 4.6%; Score 109; DB 12; Length 841;
Best Local Similarity 65.4%; Pred. No. 1.1e-15;
Matches 176; Conservative 0; Mismatches 90; Indels 3; Gaps 1;

QY 2090 GAAGATCCAAATAGACTTCAAGCGGATGAAAGCCGATGAGAGAAACCTGCTGAG 2149
DB 633 GAGTCTCCAAACCAACAGAGAGACCGAGAAAGAAATCCAGAAATCCACAGAGAA 574
QY 2150 CCAGAGTCCCTCAAGTAGAGACTGAAAGTAGAAGCCCACTCAAGAGCAGAGATT 2209

DB 573 TCAGAGAAACCTCAGGTGAGACTGAAAGTTAAGAA---AACTGAGAGGCTGAAGAT 517
QY 2210 TTGCTTGCCAAAGTAAAGGATTTGCTGAAAGCCCAATGCAACAGAACTTACGTGT 2269
DB 516 TTAAGTGGAAATCCAGATCCCAATTTATCAAGTCCCAAGAGAGACTCTACAGAGA 457
QY 2270 TTACGAATTAATTTGACTCTTCAATTTAGTAAATAGTATCAATGTCAGAGAGAGA 2329
DB 456 TTTAAATAATTAATTTCTATTGTCACCCAGAGCAAAATATATATGACAGAGCTGAA 397
QY 2330 AAATTACTGCGTGTGTTAAAGGAAGTAA 2358
DB 396 AAATATTGCGCTTTATTAAGAGAGTAA 368

RESULT 20

US-10-017-161-2179
Sequence 2179, Application US/10017161
Publication No. US2003014368A1
GENERAL INFORMATION:
APPLICANT: SUMA, MAKIYO
APPLICANT: ASAI, KIYOSHI
APPLICANT: ABURATANI, HIROYUKI
APPLICANT: ABURATANI, YUTAKA
TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
FILE REFERENCE: 08435/0152
CURRENT APPLICATION NUMBER: US/10/017,161
CURRENT FILING DATE: 2002-12-18
PRIOR APPLICATION NUMBER: JP 2001/246789
PRIOR FILING DATE: 2001-06-18
NUMBER OF SEQ ID NOS: 2430
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO: 2179
LENGTH: 1168
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: source
LOCATION: (1)..(1168)
FEATURE:
NAME/KEY: CDS
LOCATION: (201)..(968)
US-10-017-161-2179

Query Match 2.1%; Score 51.2; DB 12; Length 1168;
Best Local Similarity 48.3%; Pred. No. 0.087;
Matches 143; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 279 AAGATGAGATATTTGTAATGAGTCAAGGTCAGATATCTATCAAGTAGATGAAT 338
DB 204 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 263
QY 339 ACTATGTTTACCTTAAGGATCTGCCAGCGGATTAACGTCCTCAAAAGAGAAATCA 398
DB 264 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 323
QY 399 ATGCAAAAACAGAGCATAGTCAACATCGTGAAGGTGAATCTCAAGAAACGATG 458
DB 324 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 383
QY 459 CTGTTGCTTGCAAGCTTGCAAGAGAGCGCTTATCAAGATGATGATGATGATGATG 518
DB 384 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 443
QY 519 CTGATATATATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 574
DB 444 ATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 499

RESULT 21

US-10-029-386-25433
Sequence 25433, Application US/10029386


```

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO: 1597
; LENGTH: 487
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC010133.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
; US-09-864-761-1597

```

```

Query Match      1 9%; Score 44.6; DB 9; Length 487;
Best Local Similarity 51.8%; Pred. No. 2.1;
Matches 101; Conservative 0; Mismatches 94; Indels 0; Gaps 0;

```

```

QY 2012 CACCTGAGCAAGTCACATCTTAATGATGATGGGCAATGCCAGTGCATGTGTTA 2071
DB 166 CTCTGAGAACATGAGCATATCTGAAAAAGAGGAGAAAGAAAGAAAGATGAGG 225
QY 2072 GGCAAGAAAGACCAAGTGAAGATCCAAATAAGAACTTCAAAACGGATGAAGCCAGTA 2131
DB 226 GAGGAGGAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 285
QY 2132 GAGGAAACCTGCTGAGCCAGAGTCCCTCAAGTGAAGACTGAAAAAGTAGAGCCCA 2191
DB 286 GAACAGAGAAAGAAAGAAAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAA 345
QY 2192 CTCAGAGAGCAGAA 2206
DB 346 CAACAGAGAGAGAA 360

RESULT 28
US-10-312-841-2
; Sequence 2, Application US/10312841
; Publication No. US20030186277A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von bedeutenden genetischen Parametern innerhalb des MHC
; FILE REFERENCE: E01/1208/MO
; CURRENT APPLICATION NUMBER: US/10/312,841
; CURRENT FILING DATE: 2002-12-30
; NUMBER OF SEQ ID NOS: 2
; SEQ ID NO: 2
; LENGTH: 3673778
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
; NAME/KEY: unsure

```

```

; LOCATION: (379615)
; US-10-312-841-2
Query Match      1 9%; Score 44.6; DB 12; Length 3673778;
Best Local Similarity 48.6%; Pred. No. 2.1e+02;
Matches 122; Conservative 0; Mismatches 129; Indels 0; Gaps 0;
QY 1096 AATCTTAAATATAGACCTAAATCTCTTGTGATGACGTGACGAAAGTTGGGGA 1155
DB 1967570 AGATTATATAGATTAAAAATTTTGTATATAGTATATATATTAAGAGTGAAG 1967629
QY 1156 AGGATATGATTCGAAGAAAGGCACTCTGTTATGTCCTTTGGAAAGATTTACATC 1215
DB 1967630 AAAATGATAGAAATGAGAGAGATTATTTGTTATATATATATTTGAAGAATTATAT 1967689
QY 1216 TCAACTGTAAATATCTTGAAGAGCAAGTATCAAAAACAAGAGAGTGTTCACACATT 1275
DB 1967690 TAAAAATATATATAAAATATATATTTATTTATTTAAAAATAACGTTATTTAAAAATGGGT 1967749
QY 1276 AACTGCTAAAAAGAAATGTTGCTCTGTCGACCAAGATTTATGATTAAGCATATPA 1335
DB 1967750 AATGATATGAAAGAGATTTTACGTCGTTGTTATTTATATATATAGTGAAGTAAT 1967809
QY 1336 TCTGTTAACTG 1346
DB 1967810 TTTTAAACGG 1967820

```

```

RESULT 29
US-09-864-761-19608/C
; Sequence 19608, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687

```

```

; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 19608
; LENGTH: 766
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL008720.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89
; OTHER INFORMATION: EXPRESSED IN HELI00, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN SWISSPROT HIT: P17164, EVALUATE 4.10e+00
; OTHER INFORMATION: EST_HUMAN HIT: AW844901.1, EVALUATE 2.90e+00
US-09-864-761-19608

```

```

Query Match          1.8%; Score 43.6; DB 9; Length 766;
Best Local Similarity 46.4%; Pred. No. 4.5;
Matches 180; Conservative 0; Mismatches 204; Indels 4; Gaps 1;

```

```

Qy 186 ATGGCGACCACTATCATTTATTAACAATGGTAAGCTTCTTAAGCCCTATCATCACTGAAG 245
Db 429 ATGAAATGATGATGACCGGTATGATGAGAGATGATGAATGATGATGATGATGATGATG 370
Qy 246 AATTACTGATGAAGATCCCAACTTAAGCTTAAGAGATGATGATGATGATGATGATGATG 305
Db 369 ATTATGATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 310
Qy 306 AGGGTGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAGATGCTGCC 365
Db 309 ----TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 254
Qy 366 ACCGGATATAGCTCCGCTCAAAAAGAGAAATCAATGCAAAAACAGACATAGTCAAC 425
Db 253 ATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 194
Qy 426 ATCGGAAGTGAAGTCAACCAAGAAACGATGAGCTGTTGCTTGGCAGTTGGCAAGAC 485
Db 193 GTAGTGAAGTGAAGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 134
Qy 486 GCTATACTACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 545
Db 133 TTGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 74
Qy 546 ATGCTTATATGCTTCTCATGAGATCA 573
Db 73 AAGAGATGATGATGATGATGATGATG 46

```

RESULT 30

```

US-09-864-761-2825/c
; Sequence 2825, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecmiga-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456

```

```

; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 2825
; LENGTH: 1944
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL008720.1
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.89
; OTHER INFORMATION: EXPRESSED IN HELI00, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HEPA, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
US-09-864-761-2825

```

```

Query Match          1.8%; Score 43.6; DB 9; Length 1944;
Best Local Similarity 46.4%; Pred. No. 7.5;
Matches 180; Conservative 0; Mismatches 204; Indels 4; Gaps 1;

```

```

Qy 186 ATGGCGACCACTATCATTTATTAACAATGGTAAGCTTCTTAAGCCCTATCATCACTGAAG 245
Db 666 ATGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 607
Qy 246 AATTACTGATGAAGATCCCAACTATAAGCTAAAGATGAGATATGTTAATGAGTCA 305
Db 606 ATTATGATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 547
Qy 306 AGGGTGATATGTTATCAAGGTAGATGGAATACTATGTTTACCTTAAGATGCTGCC 365
Db 546 ----TGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 491
Qy 366 ACCGGATATAGCTCCGCTCAAAAAGAGAAATCAATGCAAAAACAGACATAGTCAAC 425
Db 490 ATGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 431
Qy 426 ATCGGAAGTGAAGTCAACCAAGAAACGATGAGCTGTTGCTTGGCAGTTGGCAAGAC 485
Db 430 GTAGTGAAGTGAAGTGAAGATGATGATGATGATGATGATGATGATGATGATGATGATG 371

```



```

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeo mica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 19488
; LENGTH: 660
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008125.9
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.5
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.8
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.8
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3
; OTHER INFORMATION: NT HIT: AF044255.1, EVALUUE 1.10e-01
; US-09-864-761-19488

Query Match      1.8%; Score 42.4; DB 9; Length 660;
Best Local Similarity 48.7%; Pred. No. 8.1;
Matches 146; Conservative 0; Mismatches 151; Indels 3; Gaps 1;

Oy      278  AAAGATGAGATATGTTAATAGAGTCAAGGATGATATGTTATCAGATGATGAGAAA 337
Db      108  AATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 167

Oy      338  TACTATGTTTACCTTAAGAGATGCTGCCACGCGGATTAACGTCCTGACAAAGAGAAATC 397
Db      168  GGTATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 227
```

```

Oy      398  AATGCAGAAAACAAGACATAGTCAACATGCGAAGG---TGGAACTCCAAGAAACGAT 454
Db      228  GATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 287

Oy      455  GGTGCTGTTGCTTGGACGTTCCGAAAGACGTTACTACATGATGATGTTATATCTTT 514
Db      288  AGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 347

Oy      515  AATGCTTCTGATATCATAGAGATGATGATGATGATGATGATGATGATGATGATGAT 574
Db      348  GACGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 407

RESULT 34
US-09-864-761-2772
; Sequence 2772, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeo mica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 2772
; LENGTH: 962
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC008125.9
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.6
```

OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.5
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 5.8
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 5
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.8
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3
US-09-864-761-2772

Query Match 1.8%; Score 42.4; DB 9; Length 962;
Best Local Similarity 48.7%; Pred. No. 9.9;
Matches 146; Conservative 0; Mismatches 151; Indels 3; Gaps 1;

QY 278 AAAGATGAGATATTGTTAATAGAGTCAGAGGTGAGTATGTTATCATCAAGTAGATGAAA 337
DB 424 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 483
QY 338 TACTATGTTACCTTAAGGATGCTGCCAGCGGATTAACGTCCTACAAAGAGAAATC 397
DB 484 GGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 543
QY 398 AATGACAAAAAAGAGCATATGTCATCTGTAAG--TGAATCTCAAGAAACGAT 454
DB 544 GATCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 603
QY 455 GGTGCTGTTGCTTGCGACGTTGCGACAGACCTTACTACAGAGATGATGATATCTTT 514
DB 604 AGTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 663
QY 515 AATGCTTCTGATATCATAGAGATCTGATGATGATGATGATGATGATGATGATGATGAT 574
DB 664 GACGATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 723

RESULT 35

US-09-815-242-4843
Sequence 4843, Application US/09815242
Patent No. US2002061569A1

GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4843
LENGTH: 4047
TYPE: DNA

ORGANISM: Staphylococcus aureus
US-09-815-242-4843

Query Match 1.8%; Score 42.4; DB 9; Length 4047;
Best Local Similarity 47.1%; Pred. No. 21;
Matches 130; Conservative 0; Mismatches 146; Indels 0; Gaps 0;

QY 1695 AAAAGTGCAGCTCAAGCTTACTATAAGAAAAAGATATCCTTACCTCCATCTCCAGACG 1754
DB 365 ATAAAGTTGAAGTTTCAACTGCGCAATATGATGAGCAAGCTTCACCAAAATTTACGAATG 424
QY 1755 CAGATGTTAAAGCAATCAACTGAGATAGTGACGACCTATTTCATCAATCGTGAAG 1814
DB 425 AAGATTTAACACTAAACAACTATATAGTAATCAAGAGCGTTACACCTGATTGCAAG 484
QY 1815 GGGAAAAAGCAATTCACCTGTTGCACTTCCATATATGTTGAGCATACAGTTGAGTTA 1874
DB 485 AGAATTAATCAGTGTGTAATGTTCAACCACTAAAGAGAAAAAAGAGATGACCA 544
QY 1875 AAAACGTAATTGATTTATTCCTCATAGGATCATTAACATTAATTAATTGCTTGGT 1934
DB 545 AAACGATCAACTACATTAATGTTAAAGTATGCTATCAAGCTAATGATGAATC 604
QY 1935 TTGATGATCAACATTAACAACTCCAAATGCTATA 1970
DB 605 TTGTTGATTAACATATGATTAATTAATTAAGAAATA 640

RESULT 36

US-09-815-242-9039
Sequence 9039, Application US/09815242
Patent No. US2002061569A1

GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
NUMBER OF SEQ ID NOS: 14110
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9039
LENGTH: 4050
TYPE: DNA
ORGANISM: Staphylococcus aureus
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(4050)
US-09-815-242-9039

Query Match 1.8%; Score 42.4; DB 9; Length 4050;
Best Local Similarity 47.1%; Pred. No. 21;


```

; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 269927
; LENGTH: 635
; TYPE: DNA
; ORGANISM: Human
; US-10-027-632-269927

Query Match
Best Local Similarity 56.2%; Pred. No. 8.9; Length 635;
Matches 77; Conservative 1; Mismatches 59; Indels 0; Gaps 0;

Qy 2071 AGGCAAGAAAGACACAGTGAATCCAAATAGAACTTCAAGCGAATGAGGCACT 2130
Db 388 AGAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 329
Qy 2131 AGAGAAACACTGCTGCTGACCGCAAGTCCCTCAAGTAGACTGAAAAAGTGAAGCCCA 2190
Db 328 AGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 269
Qy 2191 ACTCAAGAAAGCAAG 2207
Db 268 AGAAGAGAAAGAAAG 252

RESULT 40
US-09-864-761-19262/C
; Sequence 19262, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME- DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
```

```

; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonmax Sequence Listing Engine vers. 1.1
; SEQ ID NO 19262
; LENGTH: 305
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC011416.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.5
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.8
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 35
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.8
; OTHER INFORMATION: SWISSPROT HIT: P05156, EVALUE 9.40e+00
; OTHER INFORMATION: NT HIT: U67539.1, EVALUE 7.40e-01
; OTHER INFORMATION: EST_HUMAN HIT: BF129635.1, EVALUE 1.10e+00
; US-09-864-761-19262

Query Match
Best Local Similarity 52.9%; Pred. No. 6.7; Length 305;
Matches 90; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

Qy 2038 TGATGATGGGCAATGCCAGTACATGTGTTAGCAAGAAAGCACAGTGAAGATCC 2097
Db 232 TGCTTGGGGGGGTGGGGGAGAGAGACCGAGAAAGAGAGAGAAAGAAAGAAAGAA 173
Qy 2098 AAATAAGAACTTCAAGCGATGAAGACCAAGTACAGAGAAACACCTGTCAGCCAGAA 2157
Db 172 AGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 113
Qy 2158 CCTCAAGTAGAGACTGAAAAAGTGAAGCCCAATCAAGAAAGCAAG 2207
Db 112 AGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 63

RESULT 41
US-09-864-761-2534/C
; Sequence 2534, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME- DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
```

CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 2534
LENGTH: 496
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC011416.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 3.6
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.7
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.4
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.8
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 3.5
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.6
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.8
US-09-864-761-2534

Query Match 1.8%; Score 42; DB 9; Length 496;
Best Local Similarity 52.9%; Pred. No. 8.7;
Matches 90; Conservative 0; Mismatches 80; Indels 0; Gaps 0;

QY 2038 TGATGATGGGGCAATGCCAGTGCATGTGTAGCAAGAAAGCCACAGTGAAGATCC 2097
Db 484 TGGTTGGGGGGTGGGGGAGAGAGACCGAGAAAGAGAGAGAGAGAGAGAGAGAG 425
QY 2098 AATAAGAACTTCAAGCGAGTGAAGAGCCAGTAGAGAGAAACCTGCTGAGCCGAGAGT 2157
Db 424 AGAAGAAAG 365
QY 2158 CCCTCACTAGAGACTGAAAAAGTGAAGCCCACTCAAGAGAGAGAG 2207
Db 364 AGAAGAAAG 315

RESULT 42
US-10-017-161-1913
Sequence 1913, Application US/10017161
Publication No. US2003014368A1
GENERAL INFORMATION:
APPLICANT: SUMA, MAKIKO
APPLICANT: ASAI, KIYOSHI
APPLICANT: KIYAMA, YUTAKA
APPLICANT: ABURATANI, HIROYUKI
TITLE OF INVENTION: NOVEL G PROTEIN-COUPLED RECEPTORS
FILE REFERENCE: 084335/0152
CURRENT APPLICATION NUMBER: US/10/017,161
CURRENT FILING DATE: 2002-12-18
PRIOR APPLICATION NUMBER: JP 2001/246789
PRIOR FILING DATE: 2001-06-18
NUMBER OF SEQ ID NOS: 2430
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1913
LENGTH: 3532
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: source
LOCATION: (1) ..(3532)
FEATURE:
NAME/KEY: CDS
LOCATION: (201) ..(437)
FEATURE:
NAME/KEY: CDS
LOCATION: (2622) ..(3332)
US-10-017-161-1913

Query Match 1.8%; Score 42; DB 12; Length 3532;
Best Local Similarity 45.8%; Pred. No. 25;
Matches 184; Conservative 0; Mismatches 215; Indels 3; Gaps 1;

QY 199 TCATTATTAACAATGTAAGGTTCTTATGACCGTATCATCATGAGAAATTAATCATGAA 258
Db 2651 TGATGCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2710
QY 259 AGATCAAACTATTAAGCTAAAGATGAGATATTTGTAATGAGTCAAGGGTGTATATGT 318
Db 2711 TGACGGTATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2770
QY 319 TATCAAGTATGATGAGAAATTAATCTATGTTTACCTTAAGATGCTGCCACGGGATTAAGT 378
Db 2771 TGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2830
QY 379 CCGTAAGAAAGAGAAATCAATCGACA--AAACAAGAGATGTCNAATCGTGAAG 435
Db 2831 TGATCAGCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2890
QY 436 TGAACCTCAAGAAACGATGCTGCTGCTTGGCAAGTTCGCAAGAGCGTAACTAC 495
Db 2891 TGTAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2950
QY 496 AGATGATGCTTATATCTTTAATGCTTCTGATATCATAGAGATGCTGTATGCTTATAT 555
Db 2951 TGATAGTGGTGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3010
QY 556 CGTTCCTCAGAGATCATTAACATTAATCTTCAAGATGA 597
Db 3011 TGATAGTATTAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3052

RESULT 43
US-10-029-386-5350/C
Sequence 5350, Application US/10029386
Publication No. US20030194704A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.


```
RESULT 46
US-10-029-386-14986/c
; Sequence 14986, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: ABOICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 14986
; LENGTH: 413
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL138895.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
; OTHER INFORMATION: NT HIT: AJ223271.1, EVALUE 3.90e-01
; OTHER INFORMATION: SWISSPROT HIT: Q44052, EVALUE 2.00e+00
; OTHER INFORMATION: EST_HUMAN HIT: BG981138.1, EVALUE 1.00e-130
US-10-029-386-14986

Query Match
Best Local Similarity 54.6%; Score 41.6; DB 12; Length 413;
Matches 83; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 422 CAACATCGTGAAGTGGCAACTCCAGAAACGATGCTGCTTGGCCTTGGCACGTTCCGA 481
Db 211 CAACATCATGATGAGCAATCCCACTAAGGGGTGTTGCTTTCCAGGGGCTGTCTCCAT 152
Qy 482 GGAGCCTTACTACAGATGATGCTATATCTTTATATGCTTGCATATAGAGGATCT 541
Db 151 GGTGCTGTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 92
Qy 542 GGTGATGCTTATATGCTTCTCATGAGATCA 573
Db 91 TACGATGATGACGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 60

RESULT 47
US-10-029-386-1284/c
; Sequence 1284, Application US/10029386
; Publication No. US20030194704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; FILE REFERENCE: ABOICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 1284
; LENGTH: 571
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

```
OTHER INFORMATION: MAP TO AL138895.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 4.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
; OTHER INFORMATION: SWISSPROT HIT: Q44052, EVALUE 3.20e+00
; OTHER INFORMATION: EST_HUMAN HIT: BG981138.1, EVALUE 0.00e+00
US-10-029-386-1284

Query Match
Best Local Similarity 54.6%; Score 41.6; DB 12; Length 571;
Matches 83; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

Qy 422 CAACATCGTGAAGTGGCAACTCCAGAAACGATGCTGCTTGGCCTTGGCACGTTCCGA 481
Db 282 CAACATCATGATGAGCAATCCCACTAAGGGGTGTTGCTTTCCAGGGGCTGTCTCCAT 223
Qy 482 GGAGCCTTACTACAGATGATGCTATATCTTTATATGCTTGCATATAGAGGATCT 541
Db 222 GGTGCTGTGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 163
Qy 542 GGTGATGCTTATATGCTTCTCATGAGATCA 573
Db 162 TACGATGATGACGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 131
```

```
RESULT 48
US-09-864-864-233
; Sequence 233, Application US/09864864
; Patent No. US20020102679A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Secrist, Heather
; APPLICANT: Lodes, Michael J.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steve P.
; APPLICANT: Mannion, Jane
; APPLICANT: Benson, Darin R.
; APPLICANT: Carter, Darick
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.523
; CURRENT APPLICATION NUMBER: US/09/864,864
; CURRENT FILING DATE: 2001-05-23
; NUMBER OF SEQ ID NOS: 341
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 233
; LENGTH: 611
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(611)
; OTHER INFORMATION: n = A,T,C or G
US-09-864-864-233

Query Match
Best Local Similarity 51.6%; Score 41.6; DB 10; Length 611;
Matches 95; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

Qy 2080 AGACCACAGTGAAGTCAATAGAACTTCAAGCGATGAAGCCAGTAGAGAAAC 2139
Db 102 AGAAAAAGAAAGCAAGAGCTTAAACGAGAGACCAAGAGCTTAAAGAAAAAG 161
Qy 2140 ACCTGTGAGCCAGAGTCCCTCAAGTAGAGAGTGAAGAAAGTAGAAGCCCACTCAAGA 2199
```

Db 162 ACAAGCTGAATTGAAGCTCTGCTGCTTACTTAAGAGAAAAGAGAGAACTCAGACA 221
 QY 2200 ACCAGAGATTGCTTGTCCGAAAGTAACGATTCTAGTGAAGCCCAATGCACAGAAAC 2259
 Db 222 GCAAGCATTTGCTGCGCAAGAGAGAAAAGATATCCAGAAAAAGCCATTAAAGAGAAAG 281
 QY 2260 TCTA 2263
 Db 282 GCAA 285

RESULT 49

US-09-864-313
 ; Sequence 313, Application US/09864864
 ; Patent No. US20020102679A1
 ; GENERAL INFORMATION:

APPLICANT: Xu, Jiangchun
 APPLICANT: Mitcham, Jennifer L.
 APPLICANT: Harlocker, Susan L.
 APPLICANT: Dillon, David C.
 APPLICANT: Secrist, Heather J.
 APPLICANT: Lodes, Michael J.
 APPLICANT: Algate, Paul A.
 APPLICANT: Fling, Steve P.
 APPLICANT: Mannion, Jane
 APPLICANT: Benson, Darin R.
 APPLICANT: Carter, Darick
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
 FILE REFERENCE: 210121.523
 CURRENT APPLICATION NUMBER: US/09/864,864
 CURRENT FILING DATE: 2001-05-23
 NUMBER OF SEQ ID NOS: 341
 SOFTWARE: Corixa Invention Disclosure Database
 SEQ ID NO 313
 LENGTH: 1845
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-09-864-313

Query Match 1.7%; Score 41.6; DB 10; Length 1845;
 Best Local Similarity 51.6%; Pred. No. 22;
 Matches 95; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 2080 AGACCACTGAAGATCCAAATAGAACTTCAAAAGCGATGAAGAGCCAGTAGAGAAAC 2139
 Db 923 AGAAAAGAAAGCAAAAGCGAAGCTTAAACGAGAGAGAAAGCTTAAAGAAAACAAAG 982
 QY 2140 ACTTGCTAGCCAGAAATCCCTCAAGTAGAGACTGAAAAAGTAGAGCCCACTCAAGA 2199
 Db 983 ACAAGCTGAATTAGAAAGCTGCTGCTTACTTAAGAGAAAAGAGAGAGAACTCAGACA 1042
 QY 2200 AGCAGAGTTTGTCTTGCAGAAAGTAACGATTCTAGTGAAGCCCAATGCACAGAAAC 2259
 Db 1043 GCAAGCATTTGCTGCGCAAGAGAGAAAAGATATCCAGAAAAAGCCATTAAAGAGAAAG 1102
 QY 2260 TCTA 2263
 Db 1103 GCAA 1106

RESULT 50

US-09-919-497-50
 ; Sequence 50, Application US/09919497
 ; Patent No. US2002010662A1
 ; GENERAL INFORMATION:

APPLICANT: Mutter, George L.
 TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
 FILE REFERENCE: B0801/7225
 CURRENT APPLICATION NUMBER: US/09/919,497
 CURRENT FILING DATE: 2001-07-31
 PRIOR APPLICATION NUMBER: US 60/221,735
 PRIOR FILING DATE: 2000-07-31

; NUMBER OF SEQ ID NOS: 100
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 50
 ; LENGTH: 1860
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 US-09-919-497-50

Query Match 1.7%; Score 41.6; DB 10; Length 1860;
 Best Local Similarity 51.6%; Pred. No. 22;
 Matches 95; Conservative 0; Mismatches 89; Indels 0; Gaps 0;

QY 2080 AGACCACTGAAGATCCAAATAGAACTTCAAAAGCGATGAAGAGCCAGTAGAGAAAC 2139
 Db 923 AGAAAAGAAAGCAAAAGCGAAGCTTAAACGAGAGAGAGAAAGCTTAAAGAAAACAAAG 982
 QY 2140 ACTTGCTAGCCAGAAATCCCTCAAGTAGAGACTGAAAAAGTAGAGCCCACTCAAGA 2199
 Db 983 ACAAGCTGAATTAGAAAGCTGCTGCTTACTTAAGAGAAAAGAGAGAGAACTCAGACA 1042
 QY 2200 AGCAGAGTTTGTCTTGCAGAAAGTAACGATTCTAGTGAAGCCCAATGCACAGAAAC 2259
 Db 1043 GCAAGCATTTGCTGCGCAAGAGAGAAAAGATATCCAGAAAAAGCCATTAAAGAGAAAG 1102
 QY 2260 TCTA 2263
 Db 1103 GCAA 1106

Search completed: November 14, 2003, 03:48:54
 Job time : 813 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 07:49:56 ; Search time 76 Seconds
(without alignments)
1662.452 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 4165

Sequence: 1 SYELGLYQARTYKENVNRSY.....KLIALIKGSNPSSVSKEKIN 796

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

A_Geneseq_19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4163	100.0	796	19	AAW55090
2	4163	100.0	796	23	ABP54584
3	4163	100.0	802	24	ABU01598
4	4163	100.0	819	21	ABU01468
5	3218	77.3	821	21	AAH12727
6	3218	77.3	821	21	AAH12766
7	3218	77.3	821	23	AAU84026
8	3218	77.3	840	21	AAH12716
9	3218	77.3	840	23	AAU75933

10	2961	71.1	690	21	AAH12745	Streptococcus pneu
11	2961	71.1	690	23	AAU84093	Truncated variant
12	2812	67.5	826	21	AAU91939	S. pneumoniae 92 k
13	2797.5	67.2	819	24	ABU01597	S. pneumoniae type
14	2795	67.1	827	21	AAH18162	Streptococcus pneu
15	2788.5	67.0	819	21	AAH01469	Recombinant varian
16	2786	66.9	805	21	AAH12764	Streptococcus pneu
17	2772	66.6	820	21	AAH12755	Streptococcus pneu
18	2772	66.6	838	21	AAH01466	Streptococcus pneu
19	2772	66.6	839	24	ABU01418	Recombinant varian
20	2771	66.5	807	21	AAH12765	S. pneumoniae type
21	2760	66.3	811	21	AAH12763	Streptococcus pneu
22	2750.5	66.0	819	21	AAH12740	Streptococcus pneu
23	2750.5	66.0	819	21	AAH12754	Streptococcus pneu
24	2750.5	66.0	819	23	AAH84087	Truncated variant
25	2750.5	66.0	838	21	AAH12720	Streptococcus pneu
26	2750.5	66.0	838	23	AAU75934	Streptococcus pneu
27	2733	65.6	834	21	AAH12759	Streptococcus pneu
28	2728	65.5	816	21	AAH12756	Streptococcus pneu
29	2727	65.5	816	21	AAH12757	Streptococcus pneu
30	2720	65.3	811	21	AAH12760	Streptococcus pneu
31	2720	65.3	811	21	AAH12762	Streptococcus pneu
32	2714	65.2	816	21	AAH12758	Streptococcus pneu
33	2713	65.1	811	21	AAH12761	Streptococcus pneu
34	2649.5	63.6	763	19	AAW55095	S. pneumoniae 92 k
35	2649.5	63.6	763	23	ABP54589	S. pneumoniae Sp04
36	2431.5	58.4	721	20	AAU05753	Streptococcus pneu
37	2170	52.1	613	21	AAH12730	Streptococcus pneu
38	2170	52.1	613	23	AAU84029	Truncated variant
39	1898	45.6	1126	23	AAU84058	S. pneumoniae deri
40	1891.5	45.4	1365	23	AAU84057	S. pneumoniae deri
41	1888	45.3	555	21	AAH12734	Streptococcus pneu
42	1888	45.3	555	23	AAU84047	Truncated variant
43	1808	43.4	612	23	AAU84031	Truncated variant
44	1795.5	43.1	568	21	AAH12741	Streptococcus pneu
45	1795.5	43.1	568	23	AAU84088	Truncated variant
46	1795.5	43.1	1139	23	AAU84055	S. pneumoniae deri
47	1795.5	43.1	1378	23	AAU84053	S. pneumoniae deri
48	1666	40.0	334	21	AAH12728	Streptococcus pneu
49	1666	40.0	334	23	AAU84027	Truncated variant
50	1661	39.9	1238	23	AAU84056	S. pneumoniae deri
51	1651	39.7	999	23	AAU84051	S. pneumoniae deri
52	1651	39.6	929	23	AAU84052	S. pneumoniae deri
53	1649	39.6	428	21	AAH12735	Streptococcus pneu
54	1649	39.6	428	23	AAU84048	Truncated variant
55	1614.5	38.8	473	21	AAH12743	Streptococcus pneu
56	1614.5	38.8	473	23	AAU84091	Truncated variant
57	1591.5	38.2	1057	21	AAH12725	Streptococcus pneu
58	1591.5	38.2	1058	23	AAU84097	S. pneumoniae deri
59	1552	37.3	487	21	AAH12729	Streptococcus pneu
60	1552	37.3	487	23	AAU84028	Truncated variant
61	1552	37.3	487	23	AAU84090	Truncated variant
62	1247	29.9	679	23	AAU84092	Truncated variant
63	1247	29.9	780	21	AAH12744	Streptococcus pneu
64	1247	29.9	1019	21	AAH12722	Streptococcus pneu
65	1247	29.9	1019	21	AAH12721	Streptococcus pneu
66	1247	29.9	1019	23	AAU84021	Truncated variant
67	1247	29.9	1039	21	AAH12715	Streptococcus pneu
68	1247	29.9	1039	23	AAU75932	Streptococcus pneu
69	1246	29.9	1019	21	AAH12750	Streptococcus pneu
70	1246	29.9	1039	24	ABU01419	S. pneumoniae type
71	1243	29.8	1019	21	AAH12749	Streptococcus pneu
72	1238	29.7	1019	21	AAH12748	Streptococcus pneu
73	1238	29.7	1019	21	AAH12752	Streptococcus pneu
74	1238	29.7	1019	21	AAH12753	Streptococcus pneu
75	1217.5	29.2	489	21	AAH84022	Truncated variant
76	1217.5	29.2	489	23	AAU84022	Truncated variant
77	1217.5	29.2	509	21	AAH12724	Streptococcus pneu
78	1217.5	29.2	509	23	AAU84023	Truncated variant
79	1203.5	28.9	484	21	AAU84027	Truncated variant
80	1203.5	28.9	484	21	AAH81708	Streptococcus pneu
81	1203.5	28.9	485	21	AAH81538	Streptococcus pneu
82	1196.5	28.7	484	21	AAH12718	Streptococcus pneu

83	1187.5	28.5	447	19	AAW61228	Streptococcus pneu
84	1187.5	28.5	447	23	ABP54647	S. pneumoniae SP10
85	991.5	23.8	342	23	AAU84030	Truncated variant
86	991.5	23.8	913	23	AAU84059	S. pneumoniae deri
87	991.5	23.8	1152	23	AAU84054	S. pneumoniae deri
88	966.5	23.2	840	21	AAU76151	Streptococcus pneu
89	966.5	23.2	840	23	AAU76151	Streptococcus pneu
90	949	22.8	825	22	AAU00026	Streptococcus pyog
91	949	22.8	825	22	AAU00029	Streptococcus pyog
92	949	22.8	825	23	ABP26207	Streptococcus poly
93	945	22.7	823	21	AAU12746	Streptococcus pneu
94	940	22.6	824	21	AAU12747	Streptococcus pneu
95	937	22.5	822	23	ABP28904	Streptococcus poly
96	937	22.5	822	23	ABP29703	Streptococcus poly
97	934	22.4	822	22	AAU00028	Streptococcus agal
98	934	22.4	822	22	AAU00030	Streptococcus agal
99	929	22.3	822	21	AAV91283	Group B Streptococ
100	927	22.3	793	20	AAV27347	Group B Streptococ

ALIGNMENTS

RESULT 1
AAW55090 standard; Protein; 796 AA.

AAW55090 standard; Protein; 796 AA.

AAW55090;

02-OCT-1998 (first entry)

Streptococcus pneumoniae SP0036 protein.

Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;

detection; pneumonia; otitis media; meningitis.

Streptococcus pneumoniae.

Key Location/Qualifiers

Misc-difference 456

FT //label= "unknown

FT /note= "encoded by GNA"

W09818930-A2.

07-MAY-1998.

30-OCT-1997; 97WO-US19422.

31-OCT-1996; 96US-0029960.

(HUMA-) HUMAN GENOME SCT INC.

Choi GH, Hromockyj A, Johnson LS, Kunsch CA;

WPI; 1998-272224/24.

N-PSDB; AAV27351.

Nucleic acid encoding antigenic peptide(s) from Streptococcus pneumoniae - or their epitope-containing fragments, useful in protective or therapeutic vaccines, and for diagnosis

Claim 11; Page 59-60; 118pp; English.

The present sequence represents a protein from Streptococcus pneumoniae. The nucleic acid sequence encoding the Streptococcus pneumoniae protein can be useful in vaccines for inducing protective antibodies against Streptococcus pneumoniae, for treatment or prevention of infection e.g. pneumonia, otitis media or meningitis. Probes based on the nucleic acid are used to detect Streptococcus infection (by usual hybridisation or amplification methods), also for isolating Streptococcus genes or their allelic variants. The protein can be used similarly to detect specific antibodies in standard immunoassays, especially for diagnosing or

CC monitoring infections. Antibodies which bind the protein are used to

CC detect corresponding antigens, to purify the protein and for passive

CC immunisation (optionally coupled to a toxin). Vaccines are administered,

CC e.g. by injection, orally or through the skin, typically at 0.01-1000

CC (especially 10-300) mu g/ml per dose.

SQ Sequence 796 AA;

Query Match 100.0%; Score 4163; DB 19; Length 796;

Best local similarity 100.0%; Pred. No. 3.5e-299; Mismatches 0; Gaps 0;

Matches 796; Conservative 0; Indels 0; Gaps 0;

1 SYELGLYQARTVKNRVSVIDGKQATOKTENTLPDEVSKREGINAEOIVIKITDQGVY 60

1 SYELGLYQARTVKNRVSVIDGKQATOKTENTLPDEVSKREGINAEOIVIKITDQGVY 60

1 SYELGLYQARTVKNRVSVIDGKQATOKTENTLPDEVSKREGINAEOIVIKITDQGVY 60

61 SHGHHYHYNGKVPYDAIISELLMKDPNYLKDEDIYNEVKGGVYIKVDGKYVYLLKDA 120

61 SHGHHYHYNGKVPYDAIISELLMKDPNYLKDEDIYNEVKGGVYIKVDGKYVYLLKDA 120

61 SHGHHYHYNGKVPYDAIISELLMKDPNYLKDEDIYNEVKGGVYIKVDGKYVYLLKDA 120

121 AHAQNVRTKEINRQKQSHQREGCTPRNGAVALASQGRYTTDDGYIFNASDILEDT 180

121 AHAQNVRTKEINRQKQSHQREGCTPRNGAVALASQGRYTTDDGYIFNASDILEDT 180

121 AHAQNVRTKEINRQKQSHQREGCTPRNGAVALASQGRYTTDDGYIFNASDILEDT 180

181 GDAYIVPHGDHYHYI PKNELSASELAAEAFLSGRGLNSRTYRQNSDNTSRTNWPS 240

181 GDAYIVPHGDHYHYI PKNELSASELAAEAFLSGRGLNSRTYRQNSDNTSRTNWPS 240

181 GDAYIVPHGDHYHYI PKNELSASELAAEAFLSGRGLNSRTYRQNSDNTSRTNWPS 240

241 VSNPQTNTNTSNNSTNSQASQNDIDSLKQLYKPLSORHVESDGLVDPQAQITSRT 300

241 VSNPQTNTNTSNNSTNSQASQNDIDSLKQLYKPLSORHVESDGLVDPQAQITSRT 300

241 VSNPQTNTNTSNNSTNSQASQNDIDSLKQLYKPLSORHVESDGLVDPQAQITSRT 300

301 ARGVAVPHGDHYHYI PPSQMSLEBRIARII PLRYSNHWVPDRPEQSPQPTPEPSPG 360

301 ARGVAVPHGDHYHYI PPSQMSLEBRIARII PLRYSNHWVPDRPEQSPQPTPEPSPG 360

301 ARGVAVPHGDHYHYI PPSQMSLEBRIARII PLRYSNHWVPDRPEQSPQPTPEPSPG 360

361 PQPAPNLKIDNSSLVLSQVYKVEGVYFEEKGISRYFAKDLPSFTYKLNESLKSQES 420

361 PQPAPNLKIDNSSLVLSQVYKVEGVYFEEKGISRYFAKDLPSFTYKLNESLKSQES 420

361 PQPAPNLKIDNSSLVLSQVYKVEGVYFEEKGISRYFAKDLPSFTYKLNESLKSQES 420

421 VSHLTAKKENAVAPRDOEFYOKAYNLLTEAHKALFXNKGNSDFQALDKLERLNDESTN 480

421 VSHLTAKKENAVAPRDOEFYOKAYNLLTEAHKALFXNKGNSDFQALDKLERLNDESTN 480

421 VSHLTAKKENAVAPRDOEFYOKAYNLLTEAHKALFXNKGNSDFQALDKLERLNDESTN 480

481 KEKLVDDLAFIAPITPERIGKNSQIEYDEVRITAOQLADKYTTSQGYIFDEHDIISD 540

481 KEKLVDDLAFIAPITPERIGKNSQIEYDEVRITAOQLADKYTTSQGYIFDEHDIISD 540

481 KEKLVDDLAFIAPITPERIGKNSQIEYDEVRITAOQLADKYTTSQGYIFDEHDIISD 540

541 EGDAYVTPHMGSHWIGKDSISDKEKYAAQAYTEKGI LPPSPADYKANTGDSAAAIY 600

541 EGDAYVTPHMGSHWIGKDSISDKEKYAAQAYTEKGI LPPSPADYKANTGDSAAAIY 600

541 EGDAYVTPHMGSHWIGKDSISDKEKYAAQAYTEKGI LPPSPADYKANTGDSAAAIY 600

601 NRKVGKRIPLVRLPYWEHTVEYKNGNLIIPKHQYHNIFKAPFDHTTYAPNGYTLSD 660

601 NRKVGKRIPLVRLPYWEHTVEYKNGNLIIPKHQYHNIFKAPFDHTTYAPNGYTLSD 660

601 NRKVGKRIPLVRLPYWEHTVEYKNGNLIIPKHQYHNIFKAPFDHTTYAPNGYTLSD 660

661 LFATIKYVHEHPDRPHSNDGNGASHEVYCKDHSDEPNKFPADDEPVEETPAEPVVP 720

661 LFATIKYVHEHPDRPHSNDGNGASHEVYCKDHSDEPNKFPADDEPVEETPAEPVVP 720

661 LFATIKYVHEHPDRPHSNDGNGASHEVYCKDHSDEPNKFPADDEPVEETPAEPVVP 720

721 QVETEKYEAQKEAEVLLAKYTDSSLANATETLAGLRNNLTLLQIMONSIMABEKLLA 780

721 QVETEKYEAQKEAEVLLAKYTDSSLANATETLAGLRNNLTLLQIMONSIMABEKLLA 780

721 QVETEKYEAQKEAEVLLAKYTDSSLANATETLAGLRNNLTLLQIMONSIMABEKLLA 780

781 LKGSNPSVSKEKIN 796

781 LKGSNPSVSKEKIN 796

781 LKGSNPSVSKEKIN 796

RESULT 2

ABP54584

ID ABP54584 standard; Protein; 796 AA.

PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
 PR or ear infection
 PS Claim 1; SEQ ID No 2348; 56pp; English.

XX The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC Streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC AB556454. Also included are an antibody which binds one of the
 CC proteins, (treating a patient by administering the protein, DNA or
 CC antibody (in a composition), a kit comprising first and second primers,
 CC which are the nucleic acid cited above or fragments between nucleotides
 CC 8-100 of a sequence not defined in the specification, for amplifying a
 CC target sequence contained within a Streptococcus nucleic acid sequence,
 CC where the first primer is substantially complementary to the target
 CC sequence and the second primer is substantially complementary to the
 CC complement of the target sequence, and where the parts of the primers
 CC having substantial complementarity define the termini of the target
 CC sequence to be amplified, assay comprising contacting a test compound
 CC with the protein, and determining whether the test compound binds to the
 CC protein and a Streptococcus pneumoniae bacterium, where one or more
 CC genes encoding the proteins has been rendered inactive. The proteins,
 CC nucleic acid molecules, antibody and compositions are useful as
 CC medicaments for treating or preventing a disease or infection due to
 CC streptococcus bacteria, particularly S. pneumoniae, such as pneumonia,
 CC sepsis, otitis media or ear infection. They are also useful in developing
 CC vaccines, diagnostics and antibiotics. The methods are useful for
 CC identifying immunodominant proteins. The present sequence is one of
 CC the 2469 proteins expressed by the identified coding regions from the
 CC genomic sequence.

CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

CC Sequence 802 AA;

Query Match 100.0%; Score 4163; DB 24; Length 802;

Best Local Similarity 99.9%; Pred. No. 3.6e-299;

Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SEELGLYQARTYKNNRVSYIDGKATOKTENLTDPDEYSKRGINAEOIVITDOGYVT 60
 DB 7 STELGLYQARTYKNNRVSYIDGKATOKTENLTDPDEYSKRGINAEOIVITDOGYVT 66
 QY 61 SHGDHYHYNGKVPYDAIISELLMKDPNYKLMDEDIVNEVVGQYVIKVDGKYVYLKDA 120
 DB 67 SHGDHYHYNGKVPYDAIISELLMKDPNYKLMDEDIVNEVVGQYVIKVDGKYVYLKDA 126
 QY 121 AAADVNRTEETNRKQEHSGHREGTFRNDCAVALARSQGYRTTDDGYTFNASTIIDET 180
 DB 127 AAADVNRTEETNRKQEHSGHREGTFRNDCAVALARSQGYRTTDDGYTFNASTIIDET 186
 QY 181 GDAYIVPHGDHYHYI PKNELSASELAARAFISGNGINSNSTYRONSNDTSRTNWPS 240
 DB 187 GDAYIVPHGDHYHYI PKNELSASELAARAFISGNGINSNSTYRONSNDTSRTNWPS 246
 QY 241 VSNPGTTNTSNNNSNTNSQASQNSNDISLLKQYKLPISQRRHVSDDGLVPDPAITSTRT 300
 DB 247 VSNPGTTNTSNNNSNTNSQASQNSNDISLLKQYKLPISQRRHVSDDGLVPDPAITSTRT 306
 QY 301 ARGVAVPHGDHYHYI PYSQMSLEERIAIIPLYRYSNHWVPDSRPEQSPPTPEPSBG 360
 DB 307 ARGVAVPHGDHYHYI PYSQMSLEERIAIIPLYRYSNHWVPDSRPEQSPPTPEPSBG 366
 QY 361 POPAENLKIDSNSLSVQLVRKVGEGYVEEKGISRYVAPAKLPSETVNNLESKISKQES 420
 DB 367 POPAENLKIDSNSLSVQLVRKVGEGYVEEKGISRYVAPAKLPSETVNNLESKISKQES 426
 QY 421 VSHTLTAKENVAPRDOEFYDKAVNLLTEAHKALFENKGRNSDFQALDKLERLNDSTN 480

DB 427 VSHTLTAKENVAPRDOEFYDKAVNLLTEAHKALFENKGRNSDFQALDKLERLNDSTN 486
 QY 481 KEKLVDDLAFIATITPERGKPNISOIEYEDERVIAQLADKYTTSQGYTFDEHDIISD 540
 DB 487 KEKLVDDLAFIATITPERGKPNISOIEYEDERVIAQLADKYTTSQGYTFDEHDIISD 546
 QY 541 EGDAYVTPHMSHWIGKDSLSDEKEVAQAQYTEKEGILPPSPADVYKANFTGSAAIY 600
 DB 547 EGDAYVTPHMSHWIGKDSLSDEKEVAQAQYTEKEGILPPSPADVYKANFTGSAAIY 606
 QY 601 NRVGKERRIPLYRLPYMVEHTVEVKNGLIIPKHQYHNIIKPAWPDHTYKAPNGYTLTD 660
 DB 607 NRVGKERRIPLYRLPYMVEHTVEVKNGLIIPKHQYHNIIKPAWPDHTYKAPNGYTLTD 666
 QY 661 LFATIKYVEHPDRPNSNDGWSAHEVYKGDHSDPNKNFKADEPVEETPAEPVP 720
 DB 667 LFATIKYVEHPDRPNSNDGWSAHEVYKGDHSDPNKNFKADEPVEETPAEPVP 726
 QY 721 QVETEKYEAQKEAEVLLAKYTDSSLKANATETLAGLRNNLTQIMDNNSIMAEAEKLLA 780
 DB 727 QVETEKYEAQKEAEVLLAKYTDSSLKANATETLAGLRNNLTQIMDNNSIMAEAEKLLA 786
 QY 781 LKGSNPSVSKKIN 796
 DB 787 LKGSNPSVSKKIN 802

RESULT 4
 AAB01468
 ID AAB01468 standard; Protein; 819 AA.
 AC AAB01468;
 XX 20-OCT-2000 (first entry)
 DE Recombinant variant of Sp36 (Sp36A) of S. pneumoniae.
 XX
 KW Streptococcus pneumoniae; infection; vaccine; coiled coil region;
 KW histidine triad residue; sp36; antibody; otitis media;
 KW nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
 KW meningitis; lobar pneumonia.
 OS Streptococcus pneumoniae.
 XX
 FH Key Location/Qualifiers
 FT Region 63..68
 FT /label= Histidine triad residue
 FT Region 118..145
 FT /label= Coiled coil region
 FT Region 189..194
 FT /label= Histidine triad residue
 FT Region 309..314
 FT /label= Histidine triad residue
 FT Region 406..434
 FT /label= Coiled coil region
 FT Region 462..493
 FT /label= Coiled coil region
 FT Region 550..555
 FT /label= Histidine triad residue
 FT Region 634..639
 FT /label= Histidine triad residue
 FT Region 724..751
 FT /label= Coiled coil region

W0200037105-A2.
 29-JUN-2000.
 PD 21-DEC-1999; 99WO-US30390.
 PF 21-DEC-1998; 98US-0113048.
 XX
 XX (MEDI-) MEDIMUNE INC.

XX Johnson LS, Koenig S, Adamou JE;
 XX
 XX
 DR WPI: 2000-452129/39.
 DR N-PSDB; AAA47604.
 PT Vaccine useful for prophylaxis and treatment of pneumococcal infections
 PT such as otitis media, nasopharyngeal and bronchial infections,
 PT comprises Streptococcus pneumoniae proteins
 XX
 PS Claim 1, Page 61-64; 70pp; English.
 XX
 CC Although a number of proteins have been suggested as being involved
 CC in the pathogenicity of Streptococcus pneumoniae, there still remains
 CC a need to identify polypeptides having epitopes in common from
 CC various strains of S. pneumoniae in order to utilise such
 CC polypeptides in vaccines to protect against a wide variety of
 CC S. pneumoniae. New vaccine compositions are described which comprise a
 CC streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino
 CC acids in length that comprise at least one histidine triad residue
 CC (HxxHx) or a coiled-coil region, or an antibody directed against
 CC these features. The vaccine is useful in protecting against infection
 CC by Streptococcus pneumoniae. The vaccine composition comprising
 CC antibodies to is useful for passive immunization for treating
 CC pneumococcal infections which includes otitis media, nasopharyngeal
 CC and bronchial infections.
 XX
 SQ Sequence 819 AA;
 Query Match 100.0%; Score 4163; DB 21; Length 819;
 Best Local Similarity 99.9%; Pred. No. 3.7e-299;
 Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 SYELGIYQARTYKNNRVSYSIDGKQATQKTEMLTPREVSKREGINAEQIVIKITDGGYVT 60
 DB 21 SYELGIYQARTYKNNRVSYSIDGKQATQKTEMLTPREVSKREGINAEQIVIKITDGGYVT 80
 QY 61 SHGDHYYNGKVPYDAIISEELMKDPNYKLKDEDIVNEVGKGYIVKDGKYYVYLKDA 120
 DB 81 SHGDHYYNGKVPYDAIISEELMKDPNYKLKDEDIVNEVGKGYIVKDGKYYVYLKDA 140
 QY 121 AAADVRYTEEINRQKHSQHRREGTFRNDGAVALARSGRYTTDGYIFNADIIEDT 180
 DB 141 AAADVRYTEEINRQKHSQHRREGTFRNDGAVALARSGRYTTDGYIFNADIIEDT 200
 QY 181 GYAYIVPHGDHYHYIKNELSASELAAPFLSGRNLNSRTRYRQNSDNTSRTWVPS 240
 DB 201 GYAYIVPHGDHYHYIKNELSASELAAPFLSGRNLNSRTRYRQNSDNTSRTWVPS 260
 QY 241 VSNPGTNTNTNSNTNSQASQNSNDISLLKQLYKLPLSQHRVESDGLVPDPAQITTSRT 300
 DB 261 VSNPGTNTNTNSNTNSQASQNSNDISLLKQLYKLPLSQHRVESDGLVPDPAQITTSRT 320
 QY 301 AGVAVPHGDHYHYIFYSQMSLEERIRIIPLRYSNMWVPDRPEOPSPOPTPEPSPG 360
 DB 321 AGVAVPHGDHYHYIFYSQMSLEERIRIIPLRYSNMWVPDRPEOPSPOPTPEPSPG 380
 QY 361 POPAPNLKIDSNSLSVQLVRKYGEGVYBEKGISRYVPAKDI PSTVKNLESKLSKQPS 420
 DB 381 POPAPNLKIDSNSLSVQLVRKYGEGVYBEKGISRYVPAKDI PSTVKNLESKLSKQPS 440
 QY 421 VSHHTLAKKENVAPRPOEFYDKAYNLLTEAHKALFENKGRNSDFQALDKLLELNESNT 480
 DB 441 VSHHTLAKKENVAPRPOEFYDKAYNLLTEAHKALFENKGRNSDFQALDKLLELNESNT 500
 QY 481 KEKLVDDLAFAPITHPERLGRPNQSIEXTEDEVARIAQADKYTTSDEGYIPDEHIIISD 540
 DB 501 KEKLVDDLAFAPITHPERLGRPNQSIEXTEDEVARIAQADKYTTSDEGYIPDEHIIISD 560
 QY 541 EGDAYVTPHMGSHMTGKDSLSLXKQVAAQAYTKKGIILPPSPDAYKANPTDSDAAAY 600
 DB 561 EGDAYVTPHMGSHMTGKDSLSLXKQVAAQAYTKKGIILPPSPDAYKANPTDSDAAAY 620

QY 601 NRKGEKRIPLVRLPYMEHTEIVEKNGNLIIPKHDYHNIKKFWPDDHTYKANGYLTED 660
 DB 621 NRKGEKRIPLVRLPYMEHTEIVEKNGNLIIPKHDYHNIKKFWPDDHTYKANGYLTED 680
 QY 661 LFATTKYYEHPDERPHSDGNASSEHYLKKDHSDEPNKFKADEEVEETPAPEVP 720
 DB 681 LFATTKYYEHPDERPHSDGNASSEHYLKKDHSDEPNKFKADEEVEETPAPEVP 740
 QY 721 QVETKVEAKLKEAVLLAKYVTDSSLKANATELTAGLRNLTLOIMDNNSIMAEKLLA 780
 DB 741 QVETKVEAKLKEAVLLAKYVTDSSLKANATELTAGLRNLTLOIMDNNSIMAEKLLA 800
 QY 781 LKGSNPSSVSKEKIN 796
 DB 801 LKGSNPSSVSKEKIN 816
 RESULT 5
 AAB12727
 ID AAB12727 standard; Protein; 821 AA.
 XX
 AC AAB12727;
 XX
 DT 21-NOV-2000 (first entry)
 XX
 DE Streptococcus pneumoniae BVH-11M protein antigen SEQ ID NO:60.
 XX
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200039299-A2.
 XX
 PD 06-JUL-2000.
 XX
 PF 20-DEC-1999; 99WO-CA01218.
 XX
 PR 23-DEC-1998; 98US-0113800.
 XX
 PA (BIOC-) BIOCHEM PHARMA INC.
 XX
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX
 DR WPI: 2000-452397/39.
 XX
 CC Streptococcal antigens useful for vaccinating against e.g. meningitis,
 CC otitis media, bacteraemia and/or pneumonia -
 PT
 PS Claim 18; Fig 25; 106pp; English.
 XX
 CC The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents the
 CC S. pneumoniae BVH-11M protein antigen.
 XX
 SQ Sequence 821 AA;
 Query Match 77.3%; Score 3218; DB 21; Length 821;
 Best Local Similarity 75.2%; Pred. No. 3.5e-229;
 Matches 616; Conservative 65; Mismatches 104; Indels 34; Gaps 3;
 QY 1 SYELGIYQARTYKNNRVSYSIDGKQATQKTEMLTPREVSKREGINAEQIVIKITDGGYVT 60
 DB 2 AYEGLHQAYTKENNRVSYSIDGKQATQKTEMLTPREVSKREGINAEQIVIKITDGGYVT 61
 QY 61 SHGDHYYNGKVPYDAIISEELMKDPNYKLKDEDIVNEVGKGYIVKDGKYYVYLKDA 120

```

Db 62 SHGDHYHYNGKVPYDAIISELLMKDPNYQKDSIDVINEIKGVYIKVNGKYVYLLKDA 121
Qy 121 AAADVNRKKEEINRQKQESHQRECGTSPNDGAVNLABSQGYTTDDGTFINASDIIDT 180
Db 122 AAADVNRKKEEINRQKQESHQRECGTSANDGAVNLABSQGYTTDDGTFINASDIIDT 181
Qy 181 GDAIVPHGDHYHYIPKNELSASELAAPFSGGNSNSTYRONSNDTSRTNWPS 240
Db 182 GDAIVPHGDHYHYIPKNELSASELAAPFSGGNSNSTYRONSNDTSRTNWPS 241
Qy 241 VSNPGTNTNTSNTNSNTNSQASQNSNDIDSLKQLYKPLSQRHVESDGLVPDPAQITSRT 300
Db 242 VSNPGTNTNTSNTNSNTNSQASQNSNDIDSLKQLYKPLSQRHVESDGLVPDPAQITSRT 301
Qy 301 ARGVAVPHGDHYHYIPYSQMSLEERIAIIPLRIRSNHWVDSRPEQSPQPTPEPSG 360
Db 302 ARGVAVPHGNHYHFIPEQMSLEERIRAIIPLRIRSNHWVDSRPEQSPQPTPEPSG 361
Qy 361 POPAPNLK-IDSN---SSLVSQLVRKVGEGYFEEKGISRYVFAKDLPESTVKNLSEKLS 416
Db 362 POPAPNPQAPSPNPIDKLVKAVKVGDGIVFEENGYSRYIPAKNLSAETAGIDSKLA 421
Qy 417 KOESVSHTLTAKKENVAPRDOEFYDKAVNLLTEAHKALFXNKRNSDFQALDKLERND 476
Db 422 KOESLSHKLGAKKTDLPSSDREFYKAVYDLARIHQDLDLNKGQVDFEALDNLLEKLD 481
Qy 477 ESTNEXKLVDDLAFAPITPBERIGKPSQIETEDERIRIQLADKXTTSQGYTFDEHD 536
Db 482 VSSDKVKLVDDILAFAPIRHBERIGKPNQAQITTYDDEIQVAKLAGKXTTEDEGYTFDPRD 541
Qy 537 IISDEGDAYVTPHMGSHWIGKDSLDEKKNVAAQAYTEKGLPSPSPADVANKPTGSA 596
Db 542 IISDEGDAYVTPHMGSHWIKDLSSEARAAQAAYAKKGLTPSPDTHQDSGNTAEAKGA 601
Qy 597 AAIYNRVKGEKRIPLVRLPYVWEHTVEYKNGNLIIPKQDHYNIKFAFDDHTYKAPNGY 656
Db 602 EAIYNRVKAACKVPLDRMPYVNLQYTVLEVNGSLIIPHYDHYNIKFEWFEDEGLYEAAPGY 661
Qy 657 TLEDLFAITKYVVEHPDERPHSDNGMNASEHVLGKQDHSDEPNKQFAD----- 706
Db 662 TLEDLFAITKYVVEHPDERPHSDNGMNAQADTNGTEKPESEKPEKTEKPE 721
Qy 707 -----EEVETPAPEPEVPOVETEKVPAOLKEAEVLLAKVTDPSSL 746
Db 722 EETPEEKPQSEKPSPKTEPEEPESESEPEVETEKVBEKUREADDLGKIQDPII 781
Qy 747 KANATETLAGLRNNLTQIMDNNSIMAEAKLLALLKGS 785
Db 782 KSNAKETLTGLKNLFTGQDNNTIMAEAKLLALLKGS 820

```

RESULT 6.

```

AAB12766 ID AAB12766 standard; Protein; 821 AA.
AAB12766;
XX AC
XX DT 21-NOV-2000 (first entry)
XX DE Streptococcus pneumoniae strain SP64 BVH-11 protein antigen.
XX
XX KM Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX KM prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX KM otitis media; pneumonia; immunisation; bactericidal.
XX OS Streptococcus pneumoniae.
XX PN MO2000039299-A2.
XX PD 06-JUL-2000.
XX PF 20-DEC-1999; 99MO-CA01218.

```

```

XX 23-DEC-1998; 98US-0113800.
XX (BIOC-) BIOCHEM PHARMA INC.
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteraemia and/or pneumonia -
XX Disclosure; Fig 12; 106pp; English.
XX
XX The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens
XX have bactericidal activity. The nucleic acids, encoding the protein
XX antigens, may be used for the recombinant production of the proteins
XX they encode. The protein antigens may then be used as vaccines for the
XX prevention and treatment of Streptococcal infections in mammals
XX (especially humans) which result in, e.g. meningitis, otitis media,
XX bacteraemia and/or pneumonia. The present sequence represents a
XX S. pneumoniae BVH-11 protein antigen, from the present invention.
XX
XX Sequence 821 AA;
XX
XX Query Match 77.3%; Score 3218; DB 21; Length 821;
XX Best Local Similarity 75.2%; Pred. No. 3.5e-229;
XX Matches 616; Conservative 65; Mismatches 104; Indels 34; Gaps 3;
Qy 1 SYELGQARTVKKNNRVSVIDGKQATQKTNLTPDEVSKREGINAEQIVKIDQGYVT 60
Db 2 AVELGLHQAQVKKNNRVSVIDGKQATQKTNLTPDEVSKREGINAEQIVKIDQGYVT 61
Qy 61 SHGDHYHYNGKVPYDAIISELLMKDPNYKLDKEDIYNEKGVYIKVQCKYVYLLKDA 120
Db 62 SHGDHYHYNGKVPYDAIISELLMKDPNYQKDSIDVINEIKGVYIKVNGKYVYLLKDA 121
Qy 121 AAADVNRKKEEINRQKQESHQRECGTSPNDGAVNLABSQGYTTDDGTFINASDIIDT 180
Db 122 AAADVNRKKEEINRQKQESHQRECGTSANDGAVNLABSQGYTTDDGTFINASDIIDT 181
Qy 181 GDAIVPHGDHYHYIPKNELSASELAAPFSGGNSNSTYRONSNDTSRTNWPS 240
Db 182 GDAIVPHGDHYHYIPKNELSASELAAPFSGGNSNSTYRONSNDTSRTNWPS 241
Qy 241 VSNPGTNTNTSNTNSNTNSQASQNSNDIDSLKQLYKPLSQRHVESDGLVPDPAQITSRT 300
Db 242 VSNPGTNTNTSNTNSNTNSQASQNSNDIDSLKQLYKPLSQRHVESDGLVPDPAQITSRT 301
Qy 301 ARGVAVPHGDHYHYIPYSQMSLEERIAIIPLRIRSNHWVDSRPEQSPQPTPEPSG 360
Db 302 ARGVAVPHGNHYHFIPEQMSLEERIRAIIPLRIRSNHWVDSRPEQSPQPTPEPSG 361
Qy 361 POPAPNLK-IDSN---SSLVSQLVRKVGEGYFEEKGISRYVFAKDLPESTVKNLSEKLS 416
Db 362 POPAPNPQAPSPNPIDKLVKAVKVGDGIVFEENGYSRYIPAKNLSAETAGIDSKLA 421
Qy 417 KOESVSHTLTAKKENVAPRDOEFYDKAVNLLTEAHKALFXNKRNSDFQALDKLERND 476
Db 422 KOESLSHKLGAKKTDLPSSDREFYKAVYDLARIHQDLDLNKGQVDFEALDNLLEKLD 481
Qy 477 ESTNEXKLVDDLAFAPITPBERIGKPSQIETEDERIRIQLADKXTTSQGYTFDEHD 536
Db 482 VSSDKVKLVDDILAFAPIRHBERIGKPNQAQITTYDDEIQVAKLAGKXTTEDEGYTFDPRD 541
Qy 537 IISDEGDAYVTPHMGSHWIGKDSLDEKKNVAAQAYTEKGLPSPSPADVANKPTGSA 596
Db 542 IISDEGDAYVTPHMGSHWIKDLSSEARAAQAAYAKKGLTPSPDTHQDSGNTAEAKGA 601
Qy 597 AAIYNRVKGEKRIPLVRLPYVWEHTVEYKNGNLIIPKQDHYNIKFAFDDHTYKAPNGY 656
Db 602 EAIYNRVKAACKVPLDRMPYVNLQYTVLEVNGSLIIPHYDHYNIKFEWFEDEGLYEAAPGY 661

```

Oy		657TLEDLFAITKYYVEHDESPHSNDGNGNSSEHYLGKXSDENQKPFKD-----70
Dd		662 TLIEDLATYKYVHEPHNERPHSDNGSGNSDHYQRKNKGADLNQTEKSEEPQTEKPE 721
Oy		707 -----EEPVETPAEPEVPQVETEKEVAOLKEAEVLAKVTDSL 746
Dd		722 EETPREEKPKSEKPESPKPTEEPDEBSPESBPETEKEVEKELKEAEDLLKIDPPII 781
Oy		747 KANATETLAGLANNLTLOIMDNNSIAAEKALLKLSGS 785
Dd		782 KSNAKETTLGLKNLLEFGTDNNNTIWAEBKLLALUKES 820
RESULT 7		
ID	AU84026	standard, peptide, 821 AA.
XX	AU84026;	
AC		
XX		
DT	08-MAY-2002	(first entry)
XX		
DE		Truncated variant of S. pneumoniae BVH-11, BVH-11M.
KW	BVH-3; BVH-11,	vaccine; meningitis, otitis media; bacteraemia;
KW	pneumonia; streptococcal bacterial infection;	mutant; mutein.
OS	Streptococcus pneumoniae.	
OS	Synthetic.	
XX		
PN	WO200198334-A2.	
PD		
XX	27-DEC-2001.	
PF	19-JUN-2001;	2001WO-CAO0908.
XX		
PR	20-JUN-2000;	2000US-212683P.
PA	(SHLR-) SHIRE BIOCHEM INC.	
PI	Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;	
XX		
DR	WPI; 2002-122272/16.	
PT	New Streptococcus pneumoniae BVH-3 and BVH-11 variant and	
PT	epitope-bearing polypeptides, useful as vaccine components for treating	
PT	or preventing streptococcal infections such as otitis media,	
PT	meningitis, and bacteraemia -	
PS		
XX	Example 1; Page -; 113pp; English.	
CC	The invention describes an isolated polypeptide (I) with 70-90%	
CC	identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of	
CC	BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)	
CC	comprising (I) is useful for therapeutic or prophylactic treatment of	
CC	meningitis, otitis media, bacteraemia or pneumonia infection in an	
CC	individual susceptible to these disorders. (II) is also useful for	
CC	therapeutic or prophylactic treatment of any streptococcal bacterial	
CC	infection (e.g., caused by Streptococcus pneumoniae, group A	
CC	Streptococcus such as Streptococcus pyogenes, group B Streptococcus such	
CC	as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novarcidia or	
CC	Staphylococcus aureus) in an individual susceptible to the infection.	
CC	A polynucleotide (III) encoding (I) is useful in DNA immunisation	
CC	techniques. The Streptococcus polypeptides are useful in a diagnostic	
CC	test for S. pneumoniae infection. (III) is useful for designing DNA	
CC	probes for use in detecting the presence of Streptococcus in a biological	
CC	sample suspected of containing the bacteria. The DNA probes may also be	
CC	used for detecting circulating S. pneumonia nucleic acid in a sample for	
CC	diagnosing streptococcal infections. This sequence represents a truncate	
CC	of a Streptococcus pneumonia gene used to obtain antigenic peptides,	
CC	described in the method of the invention.	
CC	Note: This sequence does not appear in the specification but has	
CC	been created according to information given in the invention.	

XX	Sequence	821 AA;
XX	Query Match	77.3%; Score 3218; DB 23; Length 821;
XX	Best Local Similarity	75.2%; Pred. No. 3.5e-229;
XX	Matches	616; Conservative 65; Mismatches 104; Indels 34; Gaps 3
QY	1 SYELGLVQARIVKKNRRNSVYIDGKQATQKTENTLPDEVSKEGGINAEQIVIKITDGGYVT	60
DB	2 AYELGLHQAOQIVKKNRRNSVYIDGKQATQKTENTLPDEVSKEGGINAEQIVIKITDGGYVT	61
QY	61 SHGHYHYNKGKVPYDAIISEELMKDQNYKLDKEDIVNEVKGYYIKVDGKYYVYLKDA	120
DB	62 SHGHYHYNKGKVPYDAIISEELMKDQNYKLDKEDIVNEIYKGGYIKVNGKYYVYLKDA	121
QY	121 AAADNVRTEKEINQKQKSHQREGGTPRNDGAVALAARSQGRYTTDDGYIFNADIIEDT	180
DB	122 AAADNVRTEKEINQKQKSHQREGGTSAKDGAVALAFASQGRYTTDDGYIFNADIIEDT	181
QY	181 GDATIVPHGDHYHYIPKNELSASELAAAEAFISGSGNLSNSTYTRQNSDNTSRTWVPS	240
DB	182 GDATIVPHGDHYHYIPKNELSASELAAAEAFISGRENLSNLTYYRRQNSDNTPRTWVPS	241
QY	241 VSNNGTNTNTNNSNNTSQAOSQSDIDSLIKOLYKPLPSQHVSDGIVPQAOTTSRT	300
DB	242 VSNNGTNTNTNNSNNTSQAOSQSDIDSLIKOLYKPLPSQHVSDGILFPQAOTTSRT	301
QY	301 ARGVAVPHGDHYHYIPYSQMSLELEERIIARIIPLRYSNHWVPDSRPEQSPQTPPEPSG	360
DB	302 ARGVAVPHGNHYHYIPYQMSLELEERIIARIIPLRYSNHWVPDSRPEESPOTPPEPS	361
QY	361 POPAPNLK-IDSN--SSLSQLVARKVEGYFEEKISRYVFAKDLPSETYKNESKLS	416
DB	362 POPAPNPQAPSNPIDKELVKEAVKVGDDGYFEENGVSRYIPAKNLASETAAGIDSKLA	421
QY	417 KQESVSHLTTLAKKENVAPRDEPFYDKAANLITLAKALFYNGKRSDDPALDKLERLND	476
DB	422 KQESLSHKGAKKTDLPSSDREFFYKAYDLLARIHQDLDLNDGRQVDFEALDNLLERLKO	481
QY	477 ESTTKKELVDLLAFLAPITRPERLGKNSQIETEYDEVRVIAQLADKYTSDGYIPDEHD	536
DB	482 VSSDKVLVDILAFIAPIRPERLGKNAQIYTTDDELQVAKLAGKYTTBEDGYIFDPRD	541
QY	537 IISDEGDAYVTPHMGSHWIGKDSISDKEKVAQAQYTYKEGILPSPDADYKANPTGDSA	596
DB	542 IISDEGDAYVTPHMTSHWIKKDSISAEARAQAQYAYKEGILPSTDHQDSGNTAEAKGA	601
QY	597 AATNRYKGEKRIPLVRLPYWERTVEYKANGNLIIPHGDHYHNTIFAMFDDHTYKAPNGY	656
DB	602 EAINRYVAAKAVPLDRMPYNIQYVEYKNGSLIIPHGDHYHNTIFEWDEGLYEAPKGY	661
QY	657 TLEDLFATIKYYVEHPDERPHSNDGMSASEHVLGKQSHSEDPNKNPKYD-----	706
DB	662 TLEDLFATIKYYVEHPDERPHSNDGMSASDHVQNNKNGQADLTNOYEKSESEKPPOTEKE	721
QY	707 -----EPPVEETPAEPVPOVETEKYEAQLEAEVLLAKYVDSGL	746
DB	722 EETPREKPKQSEKPESPKPTPEEPSESESEEPQVETEKYEBEKLEAREADLLGKIDPII	781
QY	747 KAAATETLAGLRNNITLQIMDNNSIMAEAEKLLAILKKS	785
DB	782 KSNAKETITGLKNNLIFGTQDNNITIMAEAEKLLAILLKS	820
XX	RESULT 8	
XX	AA12716	
XX	ID	AA12716 standard; Protein; 840 AA.
XX	AA12716;	
XX	21-NOV-2000	(first entry)
XX	Streptococcus pneumoniae BVH-11 protein antigen SEQ ID NO:4.	


```

Qy 301 ARGVAAPHGDHYHFIPIYSQMSLEERIAIIPLRYSNHVWVDSRPEQSPQPTPEPSG 360
Cc 302 ARGVAAPHGNHNFIPYQMSLEERIAIIPLRYSNHVWVDSRPEQSPQPTPEPSG 361
Cc 361 POPAPNPK-IDSN--SSLSQLVKRGVGEVFEKGISRYVPAKDLSEVTKULESTLS 416
Cc 362 POPAPNPOPAPSNPIDEKLVKEAVRKVGCVFEENGVSRYIPAKNLSAETAGIDSKLA 421
Qy 417 KOESVSHLTAKKEVAVPRDOEFYKAVNLTLEAKALFXNKGNSDQALDKLERLND 476
Cc 422 KOESVSHLTAKKTDLPSSDRFYNKAYDLARIHQDLNDKRGVDFEALDNLERLKD 481
Qy 477 ESTNKEKLVDDLAFAPITHPERLGRKNSQIEYTEDEVRIQAOLADKYTSDGYIFDEHD 536
Cc 482 VSSDKVKLVDDLAFAPIRHPERLGRKNSQIEYTEDEVRIQAOLADKYTSDGYIFDEHD 541
Qy 537 IISDEGDAYVTPHMGSHWIGKDSLDEKVAQAAYTEKGLTPSPADYKANPTGSA 596
Cc 542 ITSDEGDAYVTPHMGSHWIKKDSLSEAEARAAQAAYTEKGLTPSTHDQSGNTEAKGA 601
Qy 597 AAIYRVKGEKRIPLVRLPYMVEHTVEYKNGNLTIIPHKDHYNIKFAMFDDHTYKAPNGY 656
Cc 602 EAIYRVKAAKVPPLDRMPYNLQYTVVEYKNGSLIIPHYDHYNIKFEMFDEGLYEAPKGY 661
Qy 657 TLEDLFATIKYVVEHPDERPHSDNGMNA 685
Cc 662 TLEDLLATVKYVVEHPNERPHSDNGFGNA 690

```

RESULT 11

AAU84093 standard; Peptide; 690 AA.

AAU84093;

08-MAY-2002 (first entry)

Truncated variant of *S. pneumoniae* BVH-11, NEM16.

BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia; pneumonia; streptococcal bacterial infection; mutant; mutein.

Streptococcus pneumoniae.

MO200198334-A2.

27-DEC-2001.

19-JUN-2001; 2001WO-CA00908.

20-JUN-2000; 2000US-212683P.

(SHIR-) SHIRE BIOCHEM INC.

Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;

WPI; 2002-122272/16.

New *Streptococcus pneumoniae* BVH-3 and BVH-11 variant and epitope-bearing polypeptides, useful as vaccine components for treating or preventing streptococcal infections such as otitis media, meningitis, and bacteraemia

Example 1; Page -, 113pp; English.

The invention describes an isolated polypeptide (I) with 70-90% identity to *Streptococcus pneumoniae* protein BVH-3, BVH-11, A vaccine of BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II) comprising (I) is useful for therapeutic or prophylactic treatment of meningitis, otitis media, bacteraemia, or pneumonia infection in an individual susceptible to these disorders. (II) is also useful for therapeutic or prophylactic treatment of any streptococcal bacterial infection (e.g., caused by *Streptococcus pneumoniae*, group A

Streptococcus such as *Streptococcus pyogenes*, group B *Streptococcus* such as *Streptococcus agalactiae*, *S. dysgalactiae*, *S. uberis*, *S. norcia* or *Staphylococcus aureus* in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The *Streptococcus* polypeptides are useful in a diagnostic test for *S. pneumoniae* infection. (III) is useful for designing DNA probes for use in detecting the presence of *Streptococcus* in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating *S. pneumoniae* nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a truncate of a *Streptococcus pneumoniae* gene used to obtain antigenic peptides, described in the method of the invention.

Note: This sequence does not appear in the specification but has been created according to information given in the invention.

Sequence 690 AA:

Query Match 71.1%; Score 2961; DB 23; Length 690;
 Best Local Similarity 80.7%; Pred. No. 2.9e-210;
 Matches 556; Conservative 54; Mismatches 75; Indels 4; Gaps 2;

```

Qy 1 SYELGLVQARTVKKNNRVSITDGKQATOKTNTLTPDEVSKREGINAEGIVIKITDQGYVT 60
Cc 2 AVELGLHQATVKKNNRVSITDGKQATOKTNTLTPDEVSKREGINAEGIVIKITDQGYVT 61
Cc 61 SHGDHYHYNGKVPYDAIIEELMKDPNYKLKDEDIENEKGVYIKVDGKYVYLLKDA 120
Cc 62 SHGDHYHYNGKVPYDAIIEELMKDPNYKLKDEDIENEKGVYIKVNGKYVYLLKDA 121
Cc 121 AHAQNVRTKEIRINQKQESHQREGGTPRNDGAVALARSQGRYTTDGYIFNASDIIDET 180
Cc 122 AHAQNVRTKEIRINQKQESHQREGGTSANDGAVALARSQGRYTTDGYIFNASDIIDET 181
Qy 181 GDAIIVPHGDHYHYIPKNELSASELAALASGRGNSRRTYRQNSDNTSRNNWPS 240
Cc 182 GDAIIVPHGDHYHYIPKNELSASELAALASGRGNSRRTYRQNSDNTSRNNWPS 241
Qy 241 VSNFGTTNTNNSNNSNTNSQASQNSNDISLKLQYKPLPSORHESDGLVDPQAQITSR 300
Cc 242 VSNFGTTNTNNSNNSNTNSQASQNSNDISLKLQYKPLPSORHESDGLVDPQAQITSR 301
Qy 301 ARGVAAPHGDHYHFIPIYSQMSLEERIAIIPLRYSNHVWVDSRPEQSPQPTPEPSG 360
Cc 302 ARGVAAPHGNHNFIPYQMSLEERIAIIPLRYSNHVWVDSRPEQSPQPTPEPSG 361
Qy 361 POPAPNPK-IDSN--SSLSQLVKRGVGEVFEKGISRYVPAKDLSEVTKULESTLS 416
Cc 362 POPAPNPOPAPSNPIDEKLVKEAVRKVGCVFEENGVSRYIPAKNLSAETAGIDSKLA 421
Qy 417 KOESVSHLTAKKEVAVPRDOEFYKAVNLTLEAKALFXNKGNSDQALDKLERLND 476
Cc 422 KOESVSHLTAKKTDLPSSDRFYNKAYDLARIHQDLNDKRGVDFEALDNLERLKD 481
Qy 477 ESTNKEKLVDDLAFAPITHPERLGRKNSQIEYTEDEVRIQAOLADKYTSDGYIFDEHD 536
Cc 482 VSSDKVKLVDDLAFAPIRHPERLGRKNSQIEYTEDEVRIQAOLADKYTSDGYIFDEHD 541
Qy 537 IISDEGDAYVTPHMGSHWIGKDSLDEKVAQAAYTEKGLTPSPADYKANPTGSA 596
Cc 542 ITSDEGDAYVTPHMGSHWIKKDSLSEAEARAAQAAYTEKGLTPSTHDQSGNTEAKGA 601
Qy 597 AAIYRVKGEKRIPLVRLPYMVEHTVEYKNGNLTIIPHKDHYNIKFAMFDDHTYKAPNGY 656
Cc 602 EAIYRVKAAKVPPLDRMPYNLQYTVVEYKNGSLIIPHYDHYNIKFEMFDEGLYEAPKGY 661
Qy 657 TLEDLFATIKYVVEHPDERPHSDNGMNA 685
Cc 662 TLEDLLATVKYVVEHPNERPHSDNGFGNA 690

```

RESULT 12
 AA91939
 ID AA91939 standard; Protein; 826 AA.

```

XX AC AAY91939;
XX DT 19-JUL-2000 (first entry)
XX DE S. pneumoniae 92 kDa human C3-degrading protein.
XX KW Human C3-degrading protein; 92 kDa; immunostimulatory; vaccine;
XX inhibitor; inflammation; organ rejection; xenotransplantation.
XX OS Streptococcus pneumoniae.
XX PN WO200017370-A1.
XX PD 30-MAR-2000.
XX PF 24-SEP-1999; 99MO-US22362.
XX PR 24-SEP-1998; 98US-0101736.
XX PR 31-MAR-1999; 99US-0283094.
XX PA (MIND) UNIV MINNESOTA.
XX PA (AMCY) AMERICAN CYANAMID CO.
XX PI Hostetter MK, Finkel DJ, Cheng Q, Green BA, Masi AW,
XX DR MPI; 2000-283594/24.
XX DR N-PSDB; AAA08557.
XX PT Isolated polypeptide is used to stimulate immune system and immunize or
XX PT treat a mammalian subject against Streptococcus pneumoniae infection or
XX PT colonization
XX PS Claim 8; Page 55-57; 63pp; English.
XX CC The present sequence, isolated from Streptococcus pneumoniae, is a
XX CC human C3-degrading protein of about 92 kDa. This sequence may encompass
XX CC a smaller, approximately 20 kDa protein (see AAY91938), also having
XX CC human C3-degrading activity. The DNA sequences (AAA08556-57) can be
XX CC used for producing an immune response to Streptococcus pneumoniae in a
XX CC mammal. Antibodies against the proteins can be used to inhibit
XX CC S. pneumoniae-mediated C3 degradation. C3-mediated inflammation and
XX CC rejection in xenotransplantation can be inhibited by expressing the
XX CC nucleic acid sequences on the surface of an organ of an animal. In
XX CC particular, the polypeptides are useful for stimulating the immune
XX CC system and are effective to immunize or treat a mammalian subject
XX CC against Streptococcus pneumoniae infection or colonization.
XX SQ Sequence 826 AA;
XX
Query Match 67.5%; Score 2812; DB 21; Length 826;
Best Local Similarity 67.4%; Pred. No. 4.1e-199;
Matches 552; Conservative 81; Mismatches 138; Indels 48; Gaps 8;
QY 1 SYELGLYQARTV-KENNRVSYIDGKQATOKTENLTPDEVSKREGINAEOIVKITDQGYV 59
DB 21 SYELGHRQAGVKKESNRVSYIDGQAGKAKENLTPDEVSKREGINAEOIVKITDQGYV 80
QY 60 TSHGDHYHYNKGVPYDAIISSELLMKDPNYKLDKEDIVNEVKGVIKVGKYYVLYKD 119
DB 81 TSHGDHYHYNKGVPYDAIISSELLMKDPNYKLDKEDIVNEVKGVIKVGKYYVLYKD 140
QY 120 AAHADNVRTKEEINRQKQESCHREGTPRNDGAVALARQSGYTTDDGIYFNASDIID 179
DB 141 AAHADNIRKEETKRQKQESHNHGGS--NDQAVVAARQSGYTTDDGIYFNASDIID 198
QY 180 TGDAYIVPGHDHYHYPKNELSASELAEEAFISGRGNLSNRYRONSNDTSRTNWP 239
DB 199 TGDAYIVPGHDHYHYPKNELSASELAEEAFISGRGNLSNRYRONSNDTSRTNWP 247
QY 240 SVSNPGTTNTNNSNTNSQASQSDISLLKQLYKPLSQRHVESDGIYFPAQITSR 299
DB 248 NPAQPLSEHNHNLTVPTYHQ-NQGENISSILRELAKPLSERHVESDGIYFPAQITSR 306

```

```

QY 300 TARGVAVPHGDHYHFIPIYSQMSLEBERIARIIPLRYSNHWVPDSRPEQSPQPTPEPSP 359
DB 307 TARGVAVPHGNHYHFIPIYQMSLEBERIARIIPLRYSNHWVPDSRPEQSPQPTPEPSP 366
QY 360 GQOPAPNPK-IDSN--SSLSQVLRKVGEGYVEEKGISRYVPAKDLPEFTVKNEISKU 415
DB 367 SQOPAPNPQAPSNPIDEKLVKAVRKVDGYVEENGVSRYIPAKDLPEFTVAGIDSKU 426
QY 416 SKQESVSHTLTAKKENVAPRDOEFYDKAVNLLTEAHKALFXNKGNSDQALDKLERLN 475
DB 427 AKQESLSHKLGAKKTDLPSSDREFFYNKAYDLARIHQDLNKGQVDEALDNLLERLK 486
QY 476 DESTNKEKLVDDLAFAPITPERLKGKNSQIEYTEDEVRJAOLADKXTTSDGYIFDGH 535
DB 487 DVPSPDKVLVDIILAFAPIRPELKGKNAQITTTDEIYQAKLAGKTTEDGYIFDR 546
QY 536 DIISDEGDAYVTPHMGSHWIGKDSLSDKEKVAQAQYTKGKILPPSPADAVKANPTGDS 595
DB 547 DITSDEGDAYVTPTHSHWIKDLSSEARAQAQYAKGKLTPTPTDHDQSGNTEAG 606
QY 596 AAATNRYKGEKRIPLVRPLPVVETVVEYKGNLIIIPKDHVHNIKFAFPDHTTYKAPNG 655
DB 607 AEAIVNRVKAARKVPLDRMPYNLYQYVEYKNSLIIIPHYDHYHNIKFEWFDGLYEAPKG 666
QY 656 YTLDELFAITIKYVVEHPDERPHSNDQMGVASEHVLCKDSDPNKNFKADE-----P 709
DB 667 YTLDELATVKKYVEHPDERPHSNDQMGVASEHVLCKDSDPNKNFKADE-----P 726
QY 710 VEET-----PAPEVPQVETEKVEAQLKEAEVLLAKYVTSGL 746
DB 727 EETPREEKQSEKPEPKTEPEPESEPEQVETEKVEKLEAREADLLGKIQDPII 786
QY 747 KANATETLAGLNNTLTQIMDNNSIMAEKILALIKS 785
DB 787 KSNAKETTLGLKNLLFGTDNNTIMAEKILALIKS 825

```

RESULT 13

ABU01597

ID ABU01597 standard; Protein; 819 AA.

XX AC ABU01597;

XX DT 11-FEB-2003 (first entry)

XX DE S. pneumoniae type 4 strain protein from coding region #1173.

XX KW Bacterial meningitis; pneumonia; sepsis; otitis media;

XX KW ear infection; antiinflammatory; antibacterial; immunostimulant;

XX KW auditory; respiratory; gene therapy; vaccine.

XX OS Streptococcus pneumoniae type 4 strain.

XX PN WO200277021-A2.

XX PD 03-OCT-2002.

XX PF 27-MAR-2002; 2002MO-1B02163.

XX PR 27-MAR-2001; 2001GB-0007658.

XX PA (CHIR-) CHIRON SPA.

XX PA (GENO-) INST GENOMIC RES.

XX PI Masignani V, Tetteelin H, Frazer C;

DR MPI; 2003-040579/03.

DR N-PSDB; ABX06885.

PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,

PT useful as medicaments for treating or preventing a disease or infection

PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media

PT or ear infection -

PS Claim 1; SEQ ID No 2346; 56pp; English.

The invention relates to a protein comprising or having at least 50% identity to any of the 2469 amino acid sequences, identified in the specification (available on a computer readable format), or its fragment, expressed from 2469 of 2489 identified DNA coding regions from the *Streptococcus pneumoniae* type 4 strain genomic sequence appearing as AB556454. Also included are an antibody which binds one of the proteins, treating a patient by administering the protein, DNA or antibody (in a composition), a kit comprising first and second primers, which are the nucleic acid cited above or fragments between nucleotides 8-100 of a sequence not defined in the specification, for amplifying a target sequence contained within a *Streptococcus* nucleic acid sequence, where the first primer is substantially complementary to the target sequence and the second primer is substantially complementary to the complement of the target sequence, and where the parts of the primers having substantial complementarity define the termini of the target sequence to be amplified, assay comprising contacting a test compound with the protein, and determining whether the test compound binds to the protein and a *Streptococcus pneumoniae* bacterium, where one or more genes encoding the proteins has been rendered inactive. The proteins, nucleic acid molecules, antibody and compositions are useful as medicaments for treating or preventing a disease or infection due to *Streptococcus* bacteria, particularly *S. pneumoniae*, such as pneumonia, sepsis, otitis media or ear infection. They are also useful in developing vaccines, diagnostics and antibiotics. The methods are useful for identifying immunodominant proteins. The present sequence is one of the 2469 proteins expressed by the identified coding regions from the genomic sequence.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIP0 at http://wipo.int/pub/published_pct_sequences.

SQ **Sequence** **819 AA;**

Query Match	67.2%	Score 2797.5	DB 24	Length 819
-------------	-------	--------------	-------	------------

Matches 546; Conservative 84; Mismatches 138; Indels 47; Gaps 7;

Oy	SYELGLOA - RVKCNENRNSYIDGKATOKTEUJLTPDEYSKREGINAEOI VIKITPOGYV 59
Db	21 SYELGLOA QDQKKSBNRVAYIIDGDOGAKEULTTDEYSKREGINAEOI VIKITPOGYV 80
Oy	60 TSHGDHYHYNGKVPYDAI I SEELLMKD BNYKJLKD EDI VNEVKG YVIVKDGXYYYLKD 119
Db	81 TSHGDHYHYNGKVPYDAI I SEELLMKD BNYQJLKD SDI VNEI KG YVIVKDGXYYYLKD 140
Oy	120 AAHMDNVTKKEINNOKOFHSOHRBEGCTPRNDGAVALLASOGGYTTDDGTIFNAASDIED 179
Db	141 AAHANIRKEIKERIKOKOERSHNHNS -- RADAAVAARAOGGYTTDDGTIFNAASDIED 197
Oy	180 TGDAYIVPHGDHYHYI PKNELSASBELAAEAFLSGRGNLSNRTYRONSNDNSTRNWWP 239
Db	198 TGDATIVPHGDHYHYI PKNELSASBELAAEA VNG-----KQGRSPSSSSYNA 246
Oy	240 SVSNPCTNTNTSNNSNTNSQASQSDNDISLQOLYKPLPSQCHVSDGLVDPDAOTISR 299
Db	247 NPAQRLSENHMLTVTPTVHO--NOGENISSLRELTAKPLSEHRVHSDGLIFPDPAOTISR 305
Oy	300 TARGA VAPRHGDHYHYI PYSOMSELEERIRI I I PLRNSHWVDSRPEOPSPOPTPEPSP 359
Db	306 TARGA VAPRHGDHYHYI PYQOMSELEERIRI I I PLRNSHWVDSRPEEPSPOPTPEPSP 365
Oy	360 GPQAPANLKI DNSSSLVSO LVRKVGEGYFEEEKIGIRIYFAKDLPEBTVNELESKLSKOE 419
Db	366 SPQAPANSPID--EKLVKAVRKVGEGYFEENGVSGRYI I PAKOLSAETAAIGDSKLAKOE 423
Oy	420 SVSHTLTAKENAVAPRDOEFYDKAYNLLTEAHKALFXNKGNSDPQALDKLERLNDSECT 479
Db	424 SLSHKLGKRTKDLPSDREFFYKAYOLLARIHODLLDNGKRGVDFALPNLIERLNDVSS 483

QY	480	NKEKUYDLLAFIAITTPRERGRKNSQIETBEDSVRIAQADKYTTSSGXFEDDEITIS	539
Db	484	DKVKJVED:ILFLAPIRPRERLGRKNAQITTYVDDDEIOVAKLJGKYTEDEGYIFDPDITIS	543
QY	540	DEGDAYVYPHGHSHWICKDSLDEKEXAAQAYTEKGLIPSPPADVYKAPPTGDSAAI	599
Db	544	DEGDYVTPPHHTSHWICKDSLSEKERRAAQYAKREKGLTPPSTPDHDSGNTAKAEAI	603
QY	600	YNRVXGEKRIPLVRLPYWVEHTVEYKNGLLIPHDYHANIKEAFEDDHITYKANPYTLE	659
Db	604	YNRVGAACKVPLDRKRPYMLQYVEYKNGSLIIPHDYHINIKFEMWDEGLYEAPKGYTLE	663
QY	660	DLFATIKYYVEHPDERPHSNDGWNASBEHVILGKHOSDPNKFKADEE-----PVEET	713
Db	664	DLLATVKKYVEHPNERPHSDNGFGNASHDVQNNKNGQADNTQTEKPSSEKQPTEKEBEET	723
QY	714	-----PAPPEYVQVETEKEVKAQLKEAEVLLAKVTDSSILKANA	750
Db	724	PREEKPOSEKESPAPKTEEPSEPSSESEPPQVETEKEVEKEKLEAREADLLGIQDPIIKSNA	783
QY	751	TETLAGLNNLLTLOIMNNSITMAEKLLALLKGS	785
Db	784	KETTLGLKNNLLFGTQDNNNTIMAEAEKLLALLKES	818

RESULT 14

ID AAY81662 standard; Protein; 827 AA.

AC AAY81662;

DT 24-MAY-2000 (first entry)

DE Streptococcus pneumoniae protein sequence ID311.

KW Streptococcus pneumoniae; vaccine; screening; protein antigen;

KW pneumococcal disease.

OS Streptococcus pneumoniae.

PN WO200006737-A2.

PD 10-FEB-2000.

PF 27-JUL-1999; 99WO-GB02451.

PR 27-JUL-1998; 98GB-0016337.

XX
XX

XX

XX
1
2
3
4
5
6
7
8
9
10

XX
2

PT pneumococcal diseases and for screening agents capable of antagonizing

XX
2

AAV81501 to AAV81679 represent specifically claimed protein sequences isolated from *Streptococcus pneumoniae*. AAV05407 to AAV05590 represent specifically claimed nucleotide sequences isolated from *S. pneumoniae*. The sequences have antibacterial and antiinflammatory properties. The protein sequences, and fragments of them, are useful as immunogens and/or antigens. The nucleotide sequences can be used in vaccines and in diagnostic assays. The proteins and nucleotides can be useful for the detection and diagnosis of *S. pneumoniae*. The protein sequences are also useful for screening an agent capable of antagonising, inhibiting or interfering with the function or expression of the proteins in which they

CC agent is useful for treatment or prophylaxis of *S. pneumoniae* infection
CC and meningitis. AAA05591 to AAA05614 represent primers used in the
CC exemplification of the present invention.

SQ Sequence 827 AA;

	Score	DB 21	Length
Query Match	67.1%	2795	827
Best Local Similarity	67.0%	Pred. No. 7.5e-198	
Matches 549; Conservative	83;	Mismatches 139;	Indels 48; Gaps 8;

QY	1	SYELSLYQA-RFVKEENRRASYIDGKATQATQTEMLTPEDEYSKRGKINAEQIVIKITDQGY	59
Db	21	SYELSRHQAGQDKESNRVAIYIDGDAQGAKAEMLTPEDEYSKRGKINAEQIVIKITDQGY	80
QY	60	TSHGDIHYHNGKVPYDAIISSELLMKDIPNYKLKDEDIINVEKGVYIVKDGKYYVYLKD	119
Db	81	TSHGDIHYHNGKVPYDAIISSELLMKDIPNYQKDSIDINEIKGVIVIKDGKYYVYLKD	140
QY	120	AAHADNVTKKEINFKQKQESHQHRBEGTFRNDGAVAAIARSQGRYTTDDGYIFNASDIIED	179
Db	141	AAHADNIRTKKEIKKQKQESHSHNH-GSGANDHAVAARAQGRYTTDDGYIFNASDIIED	198
QY	180	TGDVAIVHGHGHHYH1PKNELSASELAAAEAFPLSGNGINLSNSTRYKQNSDINSRINWVP	239
Db	199	TGDVAIVHGHGHHYH1PKNELSASELAAAEAAVNG-----KQSRPSSSSSYNA	247
QY	240	SVSNEGTINNTNSNNSNTWSQASQSNDSIDLKQIKLPLSORHVESDGLVPPAQTISR	299
Db	248	NPAQGRLEBNHNLVYTPTHQ-NQSGNISLSLRELYAKLSEHNVSDGLIPPAQTISR	306
QY	300	TARGVAVPHGDHYHFIPIYSQMSLEBRIRIIRIPLRYRSHHWVPDSHPQSPQPTPEPSP	359
Db	307	TARGVAVPHGHHYHFIPIYEQMSLEKIRIRIIRIPLRYRSHHWVPDSHPQSPQPTPEPSP	366
QY	360	GPOAPNPK-IDSN---SSLVSQIVKRVQEGVYFEFKGISRYVFAADLSSETYKMLESTL	415
Db	367	SPQRPAPNQPADSNPIDEKLVEAKRVKVDGVAFEENGVSRVYIPADOLSAETAGIDSKL	426
QY	416	SKQESVSHTLTAKKENVAPRDOEFYDKAANLLTEAKALFVXNGRNSDFQALDKLIERLN	475
Db	427	AKQESLSHKLGAKKTDLPESDREFPNKAYDLARITQDLDNKGROYDEBALDNLERLK	486
QY	476	DESTNKEKLVDLLAFIAPITHPERLKGNSQILEYDEVRIAQLADKYYTSDGYIFDEH	535
Db	487	DVSPDKVLVDIILAFIAPIRPERLKGNSQIAYTDEDIQYAKLAGKYYTTEDEGYIFDPR	546
QY	536	DIISPEGAYVTPHNHGHSHWIGKDSISDKEXYAAQAYTEKGIILPPSPDADVKANPTGDS	595
Db	547	DITSDEGAYVTPHMTSHWIKKDSISEAERAAQAQYAKKEGILTPPSTHODSGNTEAKG	606
QY	596	AAATINRKGGERILVLRPYMVEHVEHYKQCNLLIIRHGDHYNINIFANPDHDTYCAPNG	655
Db	607	AEATINRKYAAKVPIDRMPYNLYQYVEYKNSLLIIRHYDHYNIKFEWDEBELYEAPKG	666
QY	656	YTLBEDLPAITIKYYVEHPDERPHPSNDGMCVNASEHVLGKQHSDEDPNKNFPADEP-----P	709
Db	667	YTLBEDLPAITIKYYVEHPDERPHPSNDGMCVNASEHVLGKQHSDEDPNKNFPADEP-----P	726
QY	710	VEET-----PAPEVPOVETEKVEALKEAVLLAKYTDSSSL	746
Db	727	EEETPREKEPQSEKPESPKPTPEEPSEPSPEESEPPQVETEKVEKLEAEEDLGKIQDPII	786
QY	747	KANATETLAGLRNNLLTLOIMDNNSIIMAEKELLALLKGS	785
Db	787	KSNAKETLTGLKNNLLFGTDNNNTIMAEKELLALLKES	825
RESULT 15			
AAB01469			
ID	AAB01469 standard; Protein; 819 AA.		
AC	AAB01469;		
XX			

DT	20-OCT-2000	(first entry)
XX		
DE	Recombinant variant of Sp36 (Sp36B) of <i>S. pneumoniae</i>	

KM Streptococcus pneumoniae, infection; vaccine; coiled coil region;
KM histidine triad residue; Sp36; antibody; otitis media;
KM nasopharyngeal infection; bronchial infection; bronchitis; sepsis
KM meningitis; lobar pneumonia.

Streptococcus pneumoniae

PN WO200037105-A2

PD 29-JUN-2000 .

PF 21-DEC-1999; 99WO-US30390

PR 21-DEC-1998; 98US-0113048

PA (MEDI-) MEDIMUNE INC.

PI johnson LS, Koenig S, Adamou JE:

XX
DR
WPT: 2000-452129/39

DR N-PSDB; AAAA47605.
XX

PT Vaccine useful for prophylaxis and treatment of pneumococcal infections
PT such as otitis media, nasopharyngeal and bronchial infections

PT comprises Streptococcus pneumoniae proteins

PS Claim 1; Page 65-69; 70pp; English.

Although a number of proteins have been suggested as being involved in the pathogenicity of *Streptococcus pneumoniae*, there still remains a need to identify polypeptides having epitopes in common from various strains of *S. pneumoniae* in order to utilize such polypeptides in vaccines to protect against a wide variety of *S. pneumoniae*. New vaccine compositions are described which comprise a *Streptococcus pneumoniae* polypeptide (or fragments) of 80 - 680 amino acids in length that comprise at least one histidine triad residue (HxxHH) or a coiled-coil region, or an antibody directed against these features. The vaccine is useful in protecting against infection by *Streptococcus pneumoniae*. The vaccine composition comprising antibodies to is useful for passive immunization for treating pneumococcal infections which includes otitis media, nasopharyngeal and bronchial infections.

SQ Sequence 819 AA;

Query Match	67.0%;	Score 2788.5;	DB 21;	Length 819;
-------------	--------	---------------	--------	-------------

```

Desc: Local similarity 88.78; E-Val: NO: 2.2E-157;
Matches 544; Conservative 85; Mismatches 139; Indels 47; Gaps 7.

```

```

QY      1 SYELGUYQA-RTVKENNVSVYIDGKOATOKTENLTPEPVSKEGREGINAEOIVIKITDOGVY 59
Db      21 SYELGUYQAQDKKESNNVAVYIDGQAQKAENLTPEPVSKEGREGINAEOIVIKITDOGVY 80
QY      60 TSHGDHYHYNNKVPYDAIISEELIMKDPNYLKDDEIVNEVKGGYIKYDGGKYIYLYKD 119
Db      81 TSHGDHYHYNNKVPYDAIISEELIMKDPNYQLDSDIVNEIKGGYIKYNGKYYLYKD 140
QY      120 AAHADNVTRKEINOKOEHGHRGCGTFRNDGAVATARSQGRITTDGDIYFNASDIIEED 179
Db      141 AAHADNVTRKEINOKOEHGHRGCGTFRNDGAVATARSQGRITTDGDIYFNASDIIEED 197
QY      180 TGDYAVYVPRGHDYHYIIPKNELASBELAAAEFLSGRGLNSRTYPRQNSDNTSRITWYP 239
Db      198 TGDYAVYVPRGHDYHYIIPKNELASBELAAAEFLYVWG-----KQSRPSSSSSYNA 246
QY      240 SVSNPGTNTNTSNNSTNSQASQSDIDSLKOLYKLPUSQRHVESDGLVFPDQAOTISR 299
Db      247 NPAPQPLSTNNHNLVTPTYHQ-NGCENISLTLRELTVAKPLSERHVESDGLVFPDQAOTISR 305

```


AAB12755
ID AAB12755 standard; Protein; 820 AA.
XX
AC AAB12755;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae strain JNR7/87 BVH-11-2 protein antigen.
XX
KM Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KM prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KM otitis media; pneumonia; immunisation; bactericidal.
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
DR MPI; 2000-452397/39.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia -
XX
PS Disclosure; Fig 12; 106pp; English.
XX
CC The present invention describes nucleic acids (i) encoding protein
CC antigens (ii) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BVH-11-2 protein antigen, from the present invention.
XX
SQ Sequence 820 AA;
Query Match 66.6%; Score 2772; DB 21; Length 820;
Best Local Similarity 65.0%; Pred. No. 3.7e-196;
Matches 541; Conservative 91; Mismatches 144; Indels 56; Gaps 8;
QY 1 SYELGLYQARTV-KENNRVSYIDGKQATKTEMLTPEDEVSKREGINAEOIVITITOGVY 59
DB 2 STELGHQHGQVKKESNRVSIDGDAQCAENLTPDEVSKREGINAEOIVITITOGVY 61
QY 60 TSHGDHYHYNGKVPYDAIISEELMKDPNYKDKEDIVNEVGGYVIKVGKYYVYLD 119
DB 62 TSHGDHYHYNGKVPYDAIISEELMKDPNYKDKEDIVNEVGGYVIKVGKYYVYLD 121
QY 120 AAHADNVRTKEIRINROKQSHQREBGTTPRNDGAVALLASQGYTTDDGYTFNADIIED 179
DB 122 AAHADNVRTKEIRINROKQSHQREBGTTPRNDGAVALLASQGYTTDDGYTFNADIIED 179
QY 180 TGDAYIVPHGDHYHYPKNELSASELAAEAFLSGGNLSNSTYRQSDNSTRNWP 239
DB 180 TGDAYIVPHGDHYHYPKNELSASELAAEAFLSGGNLSNSTYRQSDNSTRNWP 239
QY 240 SVSNPGTNTNTSNNSNTNSQASQSDIDSLIKQLYKPLSQRHVESDGLVFPDPAITR 299
DB 229 NPAQPRLSNHNLTVPYTHQ-NQGENISLRLRELAVKPLSERHVESDGLVFPDPAITR 287
QY 300 TARGVAVPHGDHYHFLPYSQMSLEERIAIIPLRYSNHWVDSRPEQSPQPTPEPSP 359
DB 288 TARGVAVPHGDHYHFLPYSQMSLEERIAIIPLRYSNHWVDSRPEQSPQPTPEPSP 347

QY 360 GPQAPNLK-IDSN---SSLVSQLVRKYVEGYVFEKGISRYVPAKDLPSFTVKNLESKL 415
DB 348 SPQAPNPQAPASNPIDELKVEAVRKVGQGVFEENGVSRYIIPADLAEFTAGIDSKL 407
QY 416 SKQESVSHTLTAKKENAVRDOEPYDKAVNLLTEAHKLFYXNGKNSDFOALDKLIERLN 475
DB 408 AKQESLSHKLGAKKTDLPSSDREFYKAYADLLARIHQDLLDNKGRQVDPEALDNLLERLK 467
QY 476 DESTNKEKLVDDLAFIATITPERLGRKNSQIEVTEDEVRIQAOLADKTTSDGYFDEH 535
DB 468 DVPSDKVKLVDDILAFIARIRPERLGRKNAQITTTDDEIQVAKLAGKTYTDEGYIFDR 527
QY 536 DIISDEGAYVTPHMGSHWIGKDSISDKEKVAQAQYTKESGILPPSPADYKAPNGTGS 595
DB 528 DITSDEGDYVTPHMTSHWIKKDSISEAERAAQAQYAEKGLTPSTTHOSGNTAEAG 587
QY 596 AAAYNRVKGKRIPIVRLPYVEHTVEVGNLIIPIKOHYHNIFKAMPDHTYKAPNG 655
DB 588 AEAIVRVKAAKKVPLDRMPYNIQYVEVNGSLIIPHYDHYNIKFEWFDEGLYEAPKG 647
QY 656 YLLEDLFAITIKYVVEHPPDRPHSNDGMGASHEVLAGK-----KDSE----- 697
DB 648 YLLEDLFAITIKYVVEHPPDRPHSNDGMGASHEVLAGK-----KDSE----- 697
QY 698 -----DPNKNFKADEPPEETPAEPVQVETEKYEAOLKEAEVLL 738
DB 708 ESDKENHAGLNPASADNLYKPSTDTBETEBEADTTDEAEIRQEVENSVINAKIADAEALL 767
QY 739 AVYDSSLKANETLAGLRNNLTQIMDNNSIMAEAKLLALLKNSPSSV 790
DB 768 EKVTDPSIRQNMETLTGLKSSLLGTCKNNNTISAEDVSLALLKESQAPAPI 819

RESULT 18
AAB01466
ID AAB01466 standard; Protein; 838 AA.
XX
AC AAB01466;
XX
DT 20-OCT-2000 (first entry)
XX
DE Recombinant variant of Sp36 (Sp36D) of S. pneumoniae.
XX
KM Streptococcus pneumoniae; infection; vaccine; coiled coil region;
KM histidine triad residue; Sp36; antibody; otitis media;
KM nasopharyngeal infection; bronchial infection; bronchitis; sepsis;
KM meningitis; lobar pneumonia.
XX
OS Streptococcus pneumoniae.
XX
FH Key Location/Qualifiers
FT Region 64..69
FT Region /label= Histidine triad residue
FT Region 188..193
FT Region /label= Histidine triad residue
FT Region 296..301
FT Region /label= Histidine triad residue
FT Region 541..546
FT Region /label= Histidine triad residue
FT Region 625..630
FT Region /label= Histidine triad residue
PN WO200037105-A2.
XX
PD 29-JUN-2000.
XX
PF 21-DEC-1999; 99WO-US30390.
XX
PR 21-DEC-1998; 98US-0113048.
XX
PA (MED1-) MEDIMUNE INC.
XX

PI Johnson LS, Koenig S, Adamou JE;
 XX WPI: 2000-452129/39.
 DR N-PSDB; AAA47602.
 XX Vaccine useful for prophylaxis and treatment of pneumococcal infections
 PT such as otitis media, nasopharyngeal and bronchial infections,
 PT comprises Streptococcus pneumoniae proteins
 PS Claim 1; Page 54-57; 70pp; English.
 XX
 CC Although a number of proteins have been suggested as being involved
 CC in the pathogenicity of Streptococcus pneumoniae, there still remains
 CC a need to identify polypeptides having epitopes in common from
 CC various strains of S. pneumoniae in order to utilize such
 CC polypeptides in vaccines to protect against a wide variety of
 CC S. pneumoniae. New vaccine compositions are described which comprise a
 CC streptococcus pneumoniae polypeptide (or fragments) of 80 - 680 amino
 CC acids in length that comprise at least one histidine triad residue
 CC (HxxHxH) or a coiled-coil region, or an antibody directed against
 CC these features. The vaccine is useful in protecting against infection
 CC by Streptococcus pneumoniae. The vaccine composition comprising
 CC antibodies to is useful for passive immunization for treating
 CC pneumococcal infections which includes otitis media, nasopharyngeal
 CC and bronchial infections.
 XX
 XX Sequence 838 AA:
 SO
 Query Match 66.6%; Score 2772; DB 21; Length 838;
 Best Local Similarity 65.0%; Pred. No. 3.9e-196;
 Matches 541; Conservative 91; Mismatches 144; Indels 56; Gaps 8;
 QY 1 SYELGLYQARY-KENNRSVYIDGKQATQKENTLPDEVSKEEGNAQIVKIDGGYV 59
 DB 21 SYELGRHQAGVYKSSNRVSYIDGQACQAKENLPDEVSKEEGNAQIVKIDGGYV 80
 QY TSHGHVHYNGKVYDAIISEELMKDPNYKLDIEDIVNEVGSGVYIKVDGKYVYVKD 119
 DB 81 TSHGHVHYNGKVYDAIISEELMKDPNYKLDIEDIVNEVGSGVYIKVDGKYVYVKD 140
 QY 120 AAHADNVRTKEEINFQKQESHQREGTTPRNDGAVALARSQGRYTTDDGYIFNASDIIED 179
 DB 141 AAHADNIRTKKEIKQKQESHSHHGGS--NDQAVVAARAQGRYTTDDGYIFNASDIIED 198
 QY 180 TGDATVPHGHHYIIPKNELSASELAAEAFLSRGLNSNRTTRKQNSDTSTNMVP 239
 DB 199 TGDATVPHGHHYIIPKNELSASELAAEAFLSRGLNSNRTTRKQNSDTSTNMVP 247
 QY 240 SVSNGTNTNTNNSNNTNSQASQNDIDSLKQYKPLSQRHVESGLVDFPQITSR 299
 DB 248 NPAQRLSENNHLYTPTTHQ--NQGENTSSILRELKAPLSRHVESDGLIFDPQITSR 306
 QY 300 TARQAVVPHGHHYIIPYQNSLEERLARIIPLYRSNHWVDSRPQSPQPTPEPSP 359
 DB 307 TARQAVVPHGHHYIIPYQNSLEERLARIIPLYRSNHWVDSRPQSPQPTPEPSP 366
 QY 360 GPQAPNLK-IDSN---SSVLSQVRYKYGEGVFEFEKISRVPFAKDPSEVYKLESKL 415
 DB 367 SPQAPNPQAPASNIIDEKLVEARVKVGDGVFENGNSRYIPAKDISAEFLAACIDKL 426
 QY 416 SKQESVSHTLTAKKKNVAPRODFDKAVNLLTEAHKLFNKGNSNFQALDKLEELN 475
 DB 427 AKQESLHKLAKKKTDLPSDBREFYNKAYDLARHQLDLNKGQVDFEALDNLLELTK 486
 QY 476 DESTKEKLVDLLAFAPITHPERLGRKNSQIEYTEDEVRIAQLADKYTTSDDGYIFDEH 535
 DB 487 DVPSDKVTLVDLLAFAPIRHPERLGRKNSQITYTDEDEIGYAKLAGYTTDEGYIFDPR 546
 QY 536 DITSDEGAYVTPPHGSHWICKDLSDEKVAALAAQYTKKCIPLPSPDADYKANPTGDS 595
 DB 547 DITSDEGAYVTPPHGSHWICKDLSDEKVAALAAQYTKKCIPLPSPDADYKANPTGDS 606
 QY 596 AAATYNRVKGKRIPLVRLPYVWEHTVEVKGNGLIIPKHDIHNIKPFAMFDDHTYKAPNG 655

DB 607 AEAATYNRVKAQKVPDLRMPVNLQYVEVKKGSLLIIPHYDHNHNIKFEWPEGLEYEARKG 666
 QY 656 YTLDELPAFTIKYVYEHDPDERPHSDGKGNASEHVLGK-----KDSF----- 697
 DB 667 YTLDELPAFTIKYVYEHDPDERPHSDGKGNASEHVLGK-----KDSF----- 726
 QY 698 -----DKNKFPKADDEEVEETPAPEVPOYETEKVEAQLKEAEVYL 738
 DB 727 ESDKENHAGLNPAADNLKPSSTDTBEFEAEEDTTDEAEIPOYENSVINAKIDAEKLL 786
 QY 739 AKYTSSSIKANATETLAGLRNNLTQIWDNNSIWAEBEKLALIKGNSPSSV 790
 DB 787 EKVTDPSTIQVAMETLGLKSSLLGTQKDNNTISAEDVSLALIKESQAPAI 838
 RESULT 19
 ABU01418
 ID ABU01418 standard; Protein; 839 AA.
 XX AC
 XX ABU01418;
 DT 11-FEB-2003 (first entry)
 XX
 DE S. pneumoniae type 4 strain protein from coding region #993.
 XX
 KW Bacterial meningitis; pneumonia; sepsis; otitis media;
 KW ear infection; antiinflammatory; antibacterial; immunostimulant;
 KW auditory; respiratory; gene therapy; vaccine.
 XX
 OS Streptococcus pneumoniae type 4 strain.
 XX
 PN WO200277021-A2.
 XX
 PD 03-OCT-2002.
 XX
 PF 27-MAR-2002; 2002WO-1B02163.
 XX
 PR 27-MAR-2001; 2001GB-0007658.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 PI Maelgani V, Tettelin H, Fraser C;
 DR WPI: 2003-040579/03.
 DR N-PSDB; ABX06705.
 XX
 PT New proteins and nucleic acid molecules from Streptococcus pneumoniae,
 PT useful as medicaments for treating or preventing a disease or infection
 PT due to streptococcus bacteria, such as pneumonia, sepsis, otitis media
 PT or ear infection
 PT
 PT Claim 1; SEQ ID NO 1986; 56pp; English.
 PS
 PS The invention relates to a protein comprising or having at least 50%
 CC identity to any of the 2469 amino acid sequences, identified in the
 CC specification (available on a computer readable format), or its fragment,
 CC expressed from 2469 of 2489 identified DNA coding regions from the
 CC streptococcus pneumoniae type 4 strain genomic sequence appearing as
 CC AB556454. Also included are an antibody which binds one of the
 CC proteins, treating a patient by administering the protein, DNA or
 CC antibody (in a composition), a kit comprising first and second primers,
 CC which are the nucleic acid cited above or fragments between nucleotides
 CC 8-100 of a sequence not defined in the specification, for amplifying a
 CC target sequence contained within a Streptococcus nucleic acid sequence,
 CC where the first primer is substantially complementary to the target
 CC sequence and the second primer is substantially complementary to the
 CC complement of the target sequence, and where the pairs of the primers
 CC having substantial complementarity define the termini of the target
 CC sequence to be amplified, assay comprising contacting a test compound
 CC with the protein, and determining whether the test compound binds to the
 CC protein and a Streptococcus pneumoniae bacterium, where one or more

CC gene encoding the proteins has been rendered inactive. The proteins,
 CC nucleic acid molecules, antibody and compositions are useful as
 CC medicaments for treating or preventing a disease or infection due to
 CC streptococcus bacteria, particularly *S. pneumoniae*, such as pneumonia,
 CC sepsis, otitis media or ear infection. They are also useful in developing
 CC vaccines, diagnostics and antibiotics. The methods are useful for
 CC identifying immunodominant proteins. The present sequence is one of
 CC the 2469 proteins expressed by the identified coding regions from the
 CC genomic sequence.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIP0 at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 839 AA;

Query Match 66.6%; Score 2772; DB 24; Length 839;
 Best Local Similarity 65.0%; Pred. No. 3,9e-196;
 Matches 541; Conservative 91; Mismatches 144; Indels 56; Gaps 8;

QY 1 SYELGLYQARTV-KENNRSYIDGKQATOKTENLTDEVSKEKGINAQIVIKITDQGYV 59
 DB 21 STELGRHQGQVKKESNRYSIDGQAGKAEMLTDEVSKEKGINAQIVIKITDQGYV 80
 QY 60 TSHGDHYHYNGKVPYDAIISEELMKDPNYKLDIEDIVNEVGKGYVIVKDGKYYVYLYKD 119
 DB 81 TSHGDHYHYNGKVPYDAIISEELMKDPNYKLDIEDIVNEVGKGYVIVKDGKYYVYLYKD 140
 QY 120 AAHADNVRTKEELNRQKQSHQREGGTPRNDGAVALAARSQGRYTTDDGYIFNADIIED 179
 DB 141 AAHADNVRTKEELNRQKQSHQSHHGGS--NDQAVVAARAQGRYTTDDGYIFNADIIED 198
 QY 180 TGDAYIVPHGDHYHYPKNELSASELAABEALFSGHNSNSTYRQNSDNTSRNWP 239
 DB 199 TGDAYIVPHGDHYHYPKNELSASELAABEALFSGHNSNSTYRQNSDNTSRNWP 247
 QY 240 SVSNPGTTNTNTSNTSNTSOASQSDNDISLLKQLYKPLSQRHVESDGLVDPQAITSR 299
 DB 248 NPAQPRLSNNHMLTVPTVHQ--NQGENTSSLLRELVAKPLSEHVSDDGLIFPAQITSR 306
 QY 300 TARGAVAVPHGDHYHYPIYQMSGLEIRIARIIPLRYSNHWVPDSRPEQSPQPTPEPSP 359
 DB 307 TARGAVAVPHGNHFIPIYQMSGLEIRIARIIPLRYSNHWVPDSRPEQSPQPTPEPSP 366
 QY 360 GROPAPNLK-IDSN---SSLVSQLVRKVGEGYFEESKGSRYFAADLSEYTKNLESLK 415
 DB 367 SROPAPNPQAPSNPIDEKLVKAEVARKVDGIVFEENGVSRYIPAKDLAEFTAAGIDSKL 426
 QY 416 SKQESVSHLTAKKENVAPRDQEFYDKAYNLLTEAKALFXNKGKNSDFQALDKLEIRLN 475
 DB 427 AKQESLSHKLGAKTDLPPSSDREFYKAYDLARITQDILLDNKRGQVDFEALDNLIERLK 486
 QY 476 DESTNKEKLVDDLAFAPITPHERLKGNSQIEYTEDEVRIQAQADKYTSDGYIFDEH 535
 DB 487 DVPSPDVKLVDDLAFAPITPHERLKGNSQIEYTEDEVRIQAQADKYTSDGYIFDPR 546
 QY 536 DIISDEGDYVTPHMHGSHWIGKDSLDEKXVAAQAYTEKGLPRSPADVAKANTGGS 595
 DB 547 DITSDEGDYVTPHMHGSHWIKKDSLSEERAAQAYAREKGLTPRSTHODSGNTEAKG 606
 QY 596 AAATVNRVAGEKRIPLVRLPYMVEHTVEYANGNLIIPHKDYHANIIFAWFDDHTYAKAPNG 655
 DB 607 AAATVNRVAAKAVPLDRMPYNLYQYVEVYKNSGLIIIPHYDHYNIFEFMFDELYAEPKG 666
 QY 666 YLLEDLFATIKYVVEHDERPHSNDGMASEHVLK-----KHSE----- 697
 DB 667 YLLEDLFATIKYVVEHDERPHSNDGMASEHVLK-----KHSE----- 726
 QY 698 -----DPNKRFADDEPVEETFAPEVQVETEKYEAQKAEVLL 738
 DB 727 ESDKENHAGLNPASADNLYKPTDTEETEAEADTTDEAEIIPQEVSVINAKIADAELL 786
 QY 739 AKVTDSLKANATETLAGLRNNLTQIMDNNSIMAEEKLALLKKSNSPSV 790

DB 787 EKVTPSIRQNMETLTGKSSLLTGKDNNTISAEBDSLALTKESQAPI 838

RESULT 20

AAAB12765

AAAB12765 standard; Protein; 807 AA.

AAAB12765;

21-NOV-2000 (first entry)

Streptococcus pneumoniae strain SP63 BVH-11 protein antigen.

Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 otitis media; pneumonia; immunisation; bactericidal.

Streptococcus pneumoniae.

MO200039299-A2.

06-JUL-2000.

20-DEC-1999; 99WO-CA01218.

23-DEC-1998; 98US-0113800.

(BIOC-) BIOCHEM PHARMA INC.

Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

WPI; 2000-452397/39.

Streptococcal antigens useful for vaccinating against e.g. meningitis,
 otitis media, bacteraemia and/or pneumonia -

Disclosure; Fig 12; 106pp; English.

The present invention describes nucleic acids (I) encoding protein
 antigens (II) from Streptococcus pneumoniae. The protein antigens
 have bactericidal activity. The nucleic acids, encoding the protein
 antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents a
 CC *S. pneumoniae* BVH-11 protein antigen, from the present invention.

Sequence 807 AA;

Query Match 66.5%; Score 2771; DB 21; Length 807;
 Best Local Similarity 66.1%; Pred. No. 4,3e-196;
 Matches 543; Conservative 87; Mismatches 138; Indels 54; Gaps 7;

QY 1 SYELGLYQARTV-KENNRSYIDGKQATOKTENLTDEVSKEKGINAQIVIKITDQGYV 59
 DB 2 STELGRHQGQVKKESNRYSIDGQAGKAEMLTDEVSKEKGINAQIVIKITDQGYV 61
 QY 60 TSHGDHYHYNGKVPYDAIISEELMKDPNYKLDIEDIVNEVGKGYVIVKDGKYYVYLYKD 119
 DB 62 TSHGDHYHYNGKVPYDAIISEELMKDPNYKLDIEDIVNEVGKGYVIVKDGKYYVYLYKD 121
 QY 120 AAHADNVRTKEELNRQKQSHQREGGTPRNDGAVALAARSQGRYTTDDGYIFNADIIED 179
 DB 122 AAHADNVRTKEELNRQKQSHQSHHGGS--RADNAAVAAARAQGRYTTDDGYIFNADIIED 178
 QY 180 TGDAYIVPHGDHYHYPKNELSASELAABEALFSGHNSNSTYRQNSDNTSRNWP 239
 DB 179 TGDAYIVPHGNHFIPIYQMSGLEIRIARIIPLRYSNHWVPDSRPEQSPQPTPEPSP 227
 QY 240 SVSNPGTTNTNTSNTSNTSOASQSDNDISLLKQLYKPLSQRHVESDGLVDPQAITSR 299

Db 228 NPAQRLSENHNLVTPTVHQ-NOGENISLRLRELYAKPLSRHVESDGLTFDPAQITSR 286
 Qy 300 TARQAVAPHGDIHYFIPTYSQMSLEERARIITPLATRYSNHWPDSRPOSPQPTPEES- 358
 Db 287 TARQAVAPHGDIHYFIPTYSQMSLEERARIITPLATRYSNHWPDSRPOSPQPTPEES- 346
 Qy 359 -----PCGPPANLKIDSNSSLSVQLVRKVGEGYFEEKIGSRVYFADLPSETYKNDES 413
 Db 347 SPQSPAPNPQAPASNID--EKLYKEVAKKVGQGYFEEKNGSRVYIPANKLSAETAGIDS 404
 Qy 414 KLSKQESVSHLTAKENAVAPRDOEFPYDANKYLLTEAHKALFXKNGRNSDFQALDKLIER 473
 Db 405 KLSKQESVSHLTAKENAVAPRDOEFPYDANKYLLTEAHKALFXKNGRNSDFQALDKLIER 464
 Qy 474 LNDESTNKEKLVDDLLAFLAPITHTPERLGPNSQIEYDEDEVRIOALDKYTSDGYTFD 533
 Db 465 LEDVPSDKVLDLFLAPIRHPERLGPNSQIEYDEDEVRIOALDKYTSDGYTFD 524
 Qy 534 EHDIIISDEGDVAVTPHMGSHWIGKDSLSDKEKVAQAAYTEKIGILPSPADAVKANPTG 593
 Db 525 PRDITSDSGDAVTPHMGSHWIKKDSLSEARAAQAAYAKEKGLTPSTHODSGNTEA 584
 Qy 594 DSAAIYVRKGEKRIPLVRLPYMEHTVEVKNGLIIPHKHYNHNIKPAFDDHTYKAP 653
 Db 585 KGAELIYVRKAKKVPIDRMPYNYLYTEVKNGLIIPHXYHYNHNIKFEWDEGLYEAP 644
 Qy 654 NGYTEDELPAITKYVVEHPDERPHSDGNSSEHVLCKKHSEDPNKNFKAD----- 706
 Db 645 KGYTLEDLLATVYVVEHPNERPHSDGNSDGHVGRNKGQADTNOTETKPESEKPOTE 704
 Qy 707 -----EPPVETPAPEVPQVETEKVEAQLKEAEVLLAKYTD 743
 Db 705 KPEETPREBKQSKPESPKTEPEEESPEESEPQVETEKVEKLEAEADLLGKIQD 764
 Qy 744 SSKANATETTLAQLRNNLTLOIMDNNSIMAEKLLALLKGS 785
 Db 765 PIKSNAKETTLGLKNNLLFTQDNNITMAEAKLLALLKES 806
 RESULT 21
 AAB12763 ID AAB12763 standard; Protein; 811 AA.
 AC AAB12763;
 DT 21-NOV-2000 (first entry)
 XX Streptococcus pneumoniae strain RX1 BVH-11 protein antigen.
 DE Streptococcus pneumoniae: BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KM otitis media; pneumonia; immunisation; bactericidal.
 OS Streptococcus pneumoniae.
 PN MO200039299-A2.
 PD 06-JUL-2000.
 PF 20-DEC-1999; 99WO-CA01218.
 PR 23-DEC-1998; 98US-0113800.
 PA (BIOC-) BIOCHEM PHARMA INC.
 PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 DR WPI, 2000-452397/39.
 PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PS otitis media, bacteraemia and/or pneumonia -
 PS Disclosure; Fig 12; 106pp; English.

XX The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents a
 CC S. pneumoniae BVH-11 protein antigen, from the present invention.
 SQ Sequence 811 AA;
 Query Match. 66.3%; Score 2760; DB 21; Length 811;
 Best Local Similarity 65.8%; Pred. No. 2,8e-195;
 Matches 542; Conservative 87; Mismatches 141; Indels 54; Gaps 8;
 1 SYELGLYQARTV-KENNRVSYIDGKQATQKENTLPDEVSKREGINAEOIYIKITDQGY 59
 2 SYELGRHQAGQVKKESNRVSYIDGQAGKAENLTPEVSKREGINAEOIYIKITDQGY 61
 60 TSHGDHYHYNGKVPYDAIISEELMKDPYKLDIEDIVNEVKGQYIKYDGYVYVVKD 119
 62 TSHGDHYHYNGKVPYDAIISEELMKDPYKLDIEDIVNEIKGQYIKYDGYVYVVKD 121
 120 AAHADNRTKEIRROKQESQREGCTPRNDGAVALLARSQRYTDDGYTFNADIIED 179
 122 AAHADNRTKEIRROKQESQREGCTPRNDGAVALLARSQRYTDDGYTFNADIIED 178
 180 TGDYAVIPHGDIHYIIPKELISASELAAEAFLSGRNLNSRTYRQNSDNTSRVWP 239
 179 TGDYAVIPHGDIHYIIPKELISASELAAEAFLSGRNLNSRTYRQNSDNTSRVWP 227
 240 SVSNPRTNTNTSNSTNSQASNDISLKLQYKLPISQRYVESDGLVFPDAQITSR 299
 228 NPAQRLSENHNLVTPTVHQ-NOGENISLRLRELYAKPLSRHVESDGLTFDPAQITSR 286
 300 TARQAVAPHGDIHYFIPTYSQMSLEERARIITPLATRYSNHWPDSRPOSPQPTPEES- 359
 287 TARQAVAPHGDIHYFIPTYSQMSLEERARIITPLATRYSNHWPDSRPOSPQPTPEES- 346
 347 SPQSPAPNPQAPASNID--EKLYKEVAKKVGQGYFEEKNGSRVYIPANKLSAETAGIDS 406
 416 SKQESVSHLTAKENAVAPRDOEFPYDANKYLLTEAHKALFXKNGRNSDFQALDKLIER 475
 407 AKQESVSHLTAKENAVAPRDOEFPYDANKYLLTEAHKALFXKNGRNSDFQALDKLIER 466
 476 LNDESTNKEKLVDDLLAFLAPITHTPERLGPNSQIEYDEDEVRIOALDKYTSDGYTFD 535
 467 DVSDDKVKLVDDLLAFLAPIRHPERLGPNSQIEYDEDEVRIOALDKYTSDGYTFD 526
 536 DIISDEGDVAVTPHMGSHWIGKDSLSDKEKVAQAAYTEKIGILPSPADAVKANPTGDS 595
 527 DITSDEGDVAVTPHMGSHWIKKDSLSEARAAQAAYAKEKGLTPSTHODSGNTEA 586
 596 AAIYVRKGEKRIPLVRLPYMEHTVEVKNGLIIPHKHYNHNIKPAFDDHTYKAPNG 655
 587 AAIYVRKGEKRIPLVRLPYMEHTVEVKNGLIIPHKHYNHNIKPAFDDHTYKAPNG 646
 656 YTEDELPAITKYVVEHPDERPHSDGNSSEHVLCKKHSEDPNKNFKADE-----P 709
 647 YTEDELPAITKYVVEHPDERPHSDGNSSEHVLCKKHSEDPNKNFKADE-----P 706
 710 VEET-----PAPEVPQVETEKVEAQLKEAEVLLAKV 741
 707 EETPREBKQSKPESPKTEPEEESPEESEPQVETEKVEKLEAEADLLGKI 766
 742 TDSKLANATETTLAQLRNNLTLOIMDNNSIMAEKLLALLKGS 785
 767 ONPIKSNAKETTLGLKNNLLFTQDNNITMAEAKLLALLKES 810

RESULT 22
AAB12740
ID AAB12740 standard; Protein; 819 AA.
XX
XX AAB12740;
XX
XX 21-NOV-2000 (first entry)
XX
XX Streptococcus pneumoniae BVH-11-2M protein antigen SEQ ID NO:73.
DE
XX
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KM prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX
XX Streptococcus pneumoniae.
OS
XX WO200039299-A2.
XX
XX 06-JUL-2000.
XX
XX 20-DEC-1999; 99WO-CA01218.
XX
XX 23-DEC-1998; 98US-0113800.
XX
XX (BIOC-) BIOCHEM PHARMA INC.
XX
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N,
PI WPI; 2000-452397/39.
XX
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia -
XX
XX Claim 18; Fig 38; 106pp; English.
PS
XX
XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC S. pneumoniae BVH-11-2M protein antigen.
XX
XX
XX Sequence 819 AA;
SQ
Query Match 66.0%; Score 2750.5; DB 21; Length 819;
Best Local Similarity 64.5%; Pred. No. 1.5e-194;
Matches 538; Conservative 90; Mismatches 145; Indels 61; Gaps 8;
QY 1 SYELGLYQARTV-KENNRSYIDGKQATQKTENLTDEVSKEGAINAEQIVIKITDGYV 59
DB 2 SYELGHHQAGQVKKESNRVSYIDGQAKAENLTDEVSKEGAINAEQIVIKITDGYV 61
QY 60 TSHGDHYHYNGVVPYDAIISEELMKDPRYKLDKEDIVNEVGVYIKYDGYVYVYVYV 119
DB 62 TSHGDHYHYNGVVPYDAIISEELMKDPRYKLDKEDIVNEVGVYIKYDGYVYVYVYV 121
QY 120 AAHADVRTKEEINRQKQESHQREGTTPRNDGAVALLARSQGYTTDDGYTFNADIIED 179
DB 122 AAHADVIRKKEIKRQKQESHQHNHNS---RADNAVAALAAQGGYTTDDGYTFNADIIED 178
QY 180 TGDAYIVPHGDHYHYIPKXELASSELAAEAFLSGGNTSNSRTYRRQNSDNTSRTNWV 239
DB 179 TGDAYIVPHGDHYHYIPKXELASSELAAEAFLSGGNTSNSRTYRRQNSDNTSRTNWV 227
QY 240 SVSNPGTNTNTNNSNTNSQASQNDISLLKQYKPLSORHVSVDGIVPPAQITSR 299
DB 228 NPVQPLSENHNTLVPTTHQ-NQGENISLLELAKPLSEHVSVDGIFPPAQITSR 286
QY 300 TARGVAVPHGDHYHYIPYSQMSLEERIAIIPLRYSNHWVDSRPEQSPQTPPEPS- 358

DB 287 TARGVAVPHGNHYHFIPYQMSLEEKRIARIIPLRYSNHWVDSRPEQSPQTPPEPS 346
QY 359 -----PGPPAPNLKIDNSSLVSQLVRKVGSGYVEEGISRYVAKPLBETVKNLS 413
DB 347 SLQAPANPQAPASNPD--EKLKVAVRVGGYVEENGVSRYIPAKDLSAETAAGIDS 404
QY 414 KLSQESVSHLTAKENVAAPRDOEFYKAVNLTLEAHALFXNKGNSDFQALDGLLR 473
DB 405 KLAQESLSHKLGAKKTTDPPSSDREFYNKAYDLARIHQDLDDNKGROYDFEVLNLLER 464
QY 474 LNDSTNKEKLVDDLLAPLAPITPBERLQKPSQIEYTEDEVRIAQLADKYTTSGYIFD 533
DB 465 LNDVSDKVKLVDDLLAPLAPIRHBERLQKPSQAQITTYTDEIQVAKLGKTYTEGYIFD 524
QY 534 EHDIIISDEGDVAVTPHMGSHWIGDLSDEKVAQAQATKKGILPPSPDADVKANPFG 593
DB 525 PDDITSDEGDVAVTPHMTSHWIKKDSLEAERAAQAAYAKKGLTPPSTDHQDSGNTBA 584
QY 594 DSAATVNRKGEKRIPLVRLPYWVEHTVEVNGNLIIPKDHYNHNIKPAWFDHTYKAP 653
DB 585 KGAELIYNRVKAQKVPFLDRMPYNIQYVEVNGSLIIPHYDHNIKEWFDGLEYAP 644
QY 654 NGYTLDEDLFATIKYVVEHDERPHSNDGWSNASEHVLGK-----KDHSE----- 697
DB 645 KGYSLDEDLATVKKYVEHNERPHSDNGRGNASDHVRKNKADQDSKPDDEKHDEVSEPT 704
QY 698 -----DPNKNFKADEEPVEETPAEPVPOVETKEVAQLKEAV 736
DB 705 HPESEKENHAGLNPSADNLVYKPSDTDETEEBEADTTDEAEIPQVENSVINAKIADA 764
QY 737 LIAKYTDSLSKANATETLAGLRNNLTLOIMDNNSIWAEBKLLALLKGSNPSGV 790
DB 765 LLEKVTDPESIRQNAETTLTGKSSLLLTGKDNNTISAEDVSLALLESQAPAI 818
RESULT 23
AAB12754
ID AAB12754 standard; Protein; 819 AA.
XX
XX AAB12754;
XX
XX 21-NOV-2000 (first entry)
XX
XX Streptococcus pneumoniae strain Sp64 BVH-11-2 protein antigen.
DE
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KM prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
XX
XX Streptococcus pneumoniae.
OS
XX WO200039299-A2.
XX
XX 06-JUL-2000.
XX
XX 20-DEC-1999; 99WO-CA01218.
XX
XX 23-DEC-1998; 98US-0113800.
XX
XX (BIOC-) BIOCHEM PHARMA INC.
XX
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N,
PI WPI; 2000-452397/39.
XX
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia -
XX
XX Disclosure; Fig 12; 106pp; English.
XX
XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens

CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents a
 CC S. pneumoniae BVH-11-2 protein antigen, from the present invention.
 XX
 XX

Sequence 819 AA:

Query Match 66.0%; Score 2750.5; DB 21; Length 819;
 Best Local Similarity 64.5%; Pred. No. 1.5e-194;
 Matches 538; Conservative 90; Mismatches 145; Indels 61; Gaps 8;

```

QY 1 SYELGLQARTV-KENNRVSYIDGKQATOKENTLPPEVSKREGINAEQVIKTTDQGV 59
DB 2 STELRHQAGQVKKSSNRVSYIDGQAKAENLTPEVSKREGINAEQVIKTTDQGV 61
QY 60 TSHGDHYHYNGKVPYDAIISEELMKDPNYKLKDEDIVNEVGKGYIKVDGKYVYVYAKD 119
DB 62 TSHGDHYHYNGKVPYDAIISEELMKDPNYQLKDSIDVNEIKGQYIKVDGKYVYVYAKD 121
QY 120 AAHADNVRTKEINROKQESHQREBGGTPRNDGAVALARSGRYTTDGYIFNADIIED 179
DB 122 AAHADNVRTKEINROKQESHQREBGGTPRNDGAVALARSGRYTTDGYIFNADIIED 178
QY 180 TGDAYIVPHGDHYHYIPKXELASAEFLASGAGNLSNRTYRROKSDNTSTNWPV 239
DB 179 TGDAYIVPHGDHYHYIPKXELASAEFLASGAGNLSNRTYRROKSDNTSTNWPV 227
QY 240 SVSNPQTTNTNTSNNSTNSQASQSNDDLSLKOLYKLPLOSRYHESDGLVDPQAQITSR 299
DB 228 NPVOGRLSNNHLVTPTTHQ-NOGENISLLRELAKLSRHHVSEGLIDPAQITSR 286
QY 300 TARGAVAPHGDHYHYIPYSQMSLEBRIARIIPLRYRSHWVPSRPEQSPQTPPEPS- 358
DB 287 TARGAVAPHGDHYHYIPYEQMSLEBRIARIIPLRYRSHWVPSRPEQSPQTPPEPS- 346
QY 359 -PGOPAPNLKIDSSLSVSQLVRKVGKGYFEKGSRYVAFADLPETKNTES 413
DB 347 SLQAPAPNPQAPASNPID--EKVKAIVRKGVGKGYFEKGSRYVAFADLPETKNTES 404
QY 414 KLSKQESVSHLTAKKENVAPRDOEFYDAVNLLEAHKALXKNGRNSDFQALDKLER 473
DB 405 KLANQESISHKLGAKKIDLPSSDRFYKAYVLLARIHQDLDNKGROVDFEVDNLIER 464
QY 474 LNDSTNKEKLVDDLLAFLAPITPBERLQKPSQIEYTEDEVRIQALDKYTTSGYIFD 533
DB 465 LKDVSSDVKLVDDLLAFLAPITPBERLQKPSQIEYTEDEVRIQALDKYTTSGYIFD 524
QY 534 EHDITSDSGDAVYTPHMGSHWIGKDSISDEKVAQAQVTEKGIPLPSPPADVXANPTG 593
DB 525 PRDITSDSGDAVYTPHMGSHWIGKDSISDEKVAQAQVTEKGIPLPSPPADVXANPTG 584
QY 594 DSAALIVRVKGEKRIPLVRLPYMVEHTVEVKNGLIIPKQHYNIKFAWDDHTYKAP 653
DB 585 KGAELIVRVKAKKVPIDRMRYNLYTEVKNGLIIPHYHYNIKFAWDDHTYKAP 644
QY 654 NGYTLIEDLPAITKYVVEHDERPHSDNGKNSAEVILGK-----KHSE----- 697
DB 645 KGYSLIEDLATKYVVEHDERPHSDNGKNSAEVILGK-----KHSE----- 694
QY 698 -----DPNNKFKADEEPVEETPAPEPVQVEKVAQOLKEAEV 736
DB 705 HPESDEKENHAGLNSADNLKYPSTDETEEBEADDTDEAEIPOEVNSVINAKIADAE 764
QY 737 LIAKTTSSILKANATETLAGLRNNLTLOIMDNNSIMAEAKLALLKSNPSSV 790
DB 765 LLEKTTDSIRONAMETLGLKSSLLGKDNNTISAEVDSLALLKSSQAPAI 818

```

RESULT 24
 AAU84087

ID AAU84087 standard; Peptide; 819 AA.
 XX
 AC AAU84087;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE Truncated variant of S. pneumoniae BVH-11-2, BVH-
 XX
 KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mutain.
 OS Streptococcus pneumoniae.
 XX
 PN WO200198334-A2.
 XX
 PD 27-DEC-2001.
 XX
 PF 19-JUN-2001; 2001WO-CA00908.
 XX
 PR 20-JUN-2000; 2000US-212683P.
 XX
 PA (SHR-) SHIRE BIOCHEM INC.
 XX
 PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 DR WPI, 2002-122272/16.
 XX
 PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia
 XX
 PS Example 1; Page -; 113pp; English.
 XX
 CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or
 CC Streptococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a truncate
 CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
 CC described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.
 XX
 XX
 SQ Sequence 819 AA:
 Query Match 66.0%; Score 2750.5; DB 23; Length 819;
 Best Local Similarity 64.5%; Pred. No. 1.5e-194;
 Matches 538; Conservative 90; Mismatches 145; Indels 61; Gaps 8;

```

QY 1 SYELGLQARTV-KENNRVSYIDGKQATOKENTLPPEVSKREGINAEQVIKTTDQGV 59
DB 2 STELRHQAGQVKKSSNRVSYIDGQAKAENLTPEVSKREGINAEQVIKTTDQGV 61
QY 60 TSHGDHYHYNGKVPYDAIISEELMKDPNYKLKDEDIVNEVGKGYIKVDGKYVYVYAKD 119
DB 62 TSHGDHYHYNGKVPYDAIISEELMKDPNYQLKDSIDVNEIKGQYIKVDGKYVYVYAKD 121
QY 120 AAHADNVRTKEINROKQESHQREBGGTPRNDGAVALARSGRYTTDGYIFNADIIED 179

```

```

Db      122 AAADNIRTEELIKROKQESHNNHNS--RADNAVAARAQGRYTTDDGYIFNADSIIED 178
QY      180 TGDVAVIPGHGDHYHFIPIKNELSASELSAAAEAFLSGRGNLSNSRYRQNSDNTSRTPWP 239
Db      179 TGDVAVIPGHGDHYHFIPIKNELSASELSAAAEAFWNG-----KQSRPSSSSSYNA 227
QY      240 SVSNPGTTNTNTSNNSNTNSQASQSDIDSLKQLYKLPLSQRHVESDGLVPDAQITSR 299
Db      228 NPVQPRLESHNHNLVTPPTHQ--NQGENISSILRELKAPLSEHVESDGLIFDPAQITSR 286
QY      300 TARGVAVIPGHGDHYHFIPIYSQMSLEERIIIPLRYSNNHWVDSRPEQSPQPTPEPS- 358
Db      287 TARGVAVIPGHGNHYHFIPIEQMSLEEKIARIIPLRYSNNHWVDSRPEQSPQPTPEPS 346
QY      359 -----RGOPAPRLKIDSNSLSVLSQVLRKVGEGYVEEKIGISRYVPAKDLPSFTVKNLS 413
Db      347 SIOPAPNPQAPSPNPID--EKLVKEAVRKVGDDGYVEENGVSRIIPAKDLSAETAAGIDS 404
QY      414 KLSKQESVSHTLTAKKENVAPRDQEFYDKAYNLLTEAHKALFYNNKGRNSDFOALDKLLER 473
Db      405 KLAQESLSHKLGAKKTDLPSSDREFYNNKAYDLLARIHQDLLDNKGRQVDFEVLNLLER 464
QY      474 LNDESTNKEKLVDDLLAFAPITHPERLGRPNQSIETEDEVYLAQLADKYTTSDGYTFD 533
Db      465 LKDVSSDKVTLVDDIIAFLAPIRHPERLGRPNQSIETEDDEIQVAKLAGKYTTEDGYTFD 524
QY      534 EHDIIISDEGDVAVTPMHGSHWIGKDSLSDKEKVAQAAYTKEGILLPSPDADVKNPPTG 593
Db      525 PRITTSDEGDVAVTPMHGSHWIKKQSLSEAEKRAAQAAYTKEGILLPSPDADVKNPPTG 584
QY      594 DSAAAIYNNVKGKRIPLVRLPYMVEHTVEVKNGNLLIIPKDHYNHIKFAFPDHYTKAP 653
Db      585 KGAEAIIYNNVKAQKAPPLDMRPYNLQYTVKNGSLIIPHYDHYNHIKFEWPEDEGYEAR 644
QY      654 NGYTTEDLPATIKYYVEHDPERPHSNDGKGNASEHVLGK-----KHSE----- 697
Db      645 KGYSLIEDLATVYKYYVEHNERPHSDNGFGNASDHYRKAKADODSKPEDDEKHEDEVSEPT 704
QY      698 -----DGNKQPKADEEPEVEETPAPEVPOYTEKVEAQLEAEV 736
Db      705 HPSDEKENHAGLNPSADNLKPSSTDEETEERAEETTDPAEIPQVENSVINAKIADAE 764
QY      737 LIAKVTDSSIKANATETLAGLRNNLTQIMDNNSINMAEAKLLALLKGSNPSSV 790
Db      765 LIEKVTDPSIRQAMETLTGLKSSLLGTQDNNTISAEVDSLALLKESQPAPI 818

RESULT 25
AAB12720
ID      AAB12720 standard; Protein; 838 AA.
XX
AC      AAB12720;
XX
DT      21-NOV-2000 (first entry)
XX
DE      Streptococcus pneumoniae BVH-11-2 protein antigen SPQ ID NO:14.
XX
KW      Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW      Propylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW      Otitis media; pneumonia; immunisation; bactericidal.
XX
OS      Streptococcus pneumoniae.
XX
PN      WO200039299-A2.
XX
PD      06-JUL-2000.
XX
PF      20-DEC-1999; 99MO-CA01218.
XX
PR      23-DEC-1998; 98US-0113800.
XX
PA      (BIOC-) BIOCHEM PHARMA INC.
XX

```

```

PI      Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX      WPI; 2000-452397/39.
DR      N-PDSB; AAA65737.
XX
PT      Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX      Otitis media, bacteraemia and/or pneumonia -
XX      Claim 18; Fig 17; 106pp; English.
XX
CC      The present invention describes nucleic acid (1) encoding protein
CC      antigens (11) from Streptococcus pneumoniae. The protein antigens
CC      have bactericidal activity. The nucleic acids, encoding the protein
CC      antigens, may be used for the recombinant production of the proteins
CC      they encode. The protein antigens may than be used as vaccines for the
CC      prevention and treatment of Streptococcal infections in mammals
CC      (especially humans) which result in, e.g. meningitis, Otitis media,
CC      bacteraemia and/or pneumonia. The present sequence represents the
CC      S. pneumoniae BVH-11-2 protein antigen.
XX
SQ      Sequence 838 AA;
Query Match 66.0%; Score 2750.5; DB 21; Length 838;
Best Local Similarity 64.5%; Pred. No. 1,5e-194;
Matches 538; Conservative 90; Mismatches 145; Indels 61; Gaps 8;
QY      1 SYELGIYQARVY-KENNRYSYIDGKATQKTENTLPDEVSKEGIAEQIVIKITQGYV 59
Db      21 SYELGHHQGVYKESNRVSYIDGDAQGAENLTPDEVSKREGINAEOIVIKITQGYV 80
QY      60 TSHGDHYHYNGKVPYDAIISSEILMKDPNYKLKDEDIVNEVYGVYIKVDGKYVYLYKD 119
Db      81 TSHGDHYHYNGKVPYDAIISSEILMKDPNYQKSDIYNELKGVIVIKVDGKYVYLYKD 140
QY      120 AAHADNVRTKEIINROKQESHQREGTPEANDGAVALASQGRYTTDDGYIFNADSIIED 179
Db      141 AAHADNIRTEELIKROKQESHNNHNS--RADNAVAARAQGRYTTDDGYIFNADSIIED 197
QY      180 TGDVAVIPGHGDHYHFIPIKNELSASELSAAAEAFLSGRGNLSNSRYRQNSDNTSRTPWP 239
Db      198 TGDVAVIPGHGDHYHFIPIKNELSASELSAAAEAFWNG-----KQSRPSSSSSYNA 246
QY      240 SVSNPGTTNTNTSNNSNTNSQASQSDIDSLKQLYKLPLSQRHVESDGLVPDAQITSR 299
Db      247 NPVQPRLESHNHNLVTPPTHQ--NQGENISSILRELKAPLSEHVESDGLIFDPAQITSR 305
QY      300 TARGVAVIPGHGDHYHFIPIYSQMSLEERIIIPLRYSNNHWVDSRPEQSPQPTPEPS- 358
Db      306 TARGVAVIPGHGNHYHFIPIEQMSLEEKIARIIPLRYSNNHWVDSRPEQSPQPTPEPS 365
QY      359 -----RGOPAPRLKIDSNSLSVLSQVLRKVGEGYVEEKIGISRYVPAKDLPSFTVKNLS 413
Db      366 SIOPAPNPQAPSPNPID--EKLVKEAVRKVGDDGYVEENGVSRIIPAKDLSAETAAGIDS 423
QY      414 KLSKQESVSHTLTAKKENVAPRDQEFYDKAYNLLTEAHKALFYNNKGRNSDFOALDKLLER 473
Db      424 KLAQESLSHKLGAKKTDLPSSDREFYNNKAYDLLARIHQDLLDNKGRQVDFEVLNLLER 483
QY      474 LNDESTNKEKLVDDLLAFAPITHPERLGRPNQSIETEDEVYLAQLADKYTTSDGYTFD 533
Db      484 LKDVSSDKVTLVDDIIAFLAPIRHPERLGRPNQSIETEDDEIQVAKLAGKYTTEDGYTFD 543
QY      534 EHDIIISDEGDVAVTPMHGSHWIGKDSLSDKEKVAQAAYTKEGILLPSPDADVKNPPTG 593
Db      544 PRITTSDEGDVAVTPMHGSHWIKKQSLSEAEKRAAQAAYTKEGILLPSPDADVKNPPTG 603
QY      594 DSAAAIYNNVKGKRIPLVRLPYMVEHTVEVKNGNLLIIPKDHYNHIKFAFPDHYTKAP 653
Db      604 KGAEAIIYNNVKAQKAPPLDMRPYNLQYTVKNGSLIIPHYDHYNHIKFEWPEDEGYEAR 663
QY      654 NGYTTEDLPATIKYYVEHDPERPHSNDGKGNASEHVLGK-----KHSE----- 697
Db      664 KGYSLIEDLATVYKYYVEHNERPHSDNGFGNASDHYRKAKADODSKPEDDEKHEDEVSEPT 723

```

QY 698 -----DPKNFKADEEPEVETPAEPEVPOVETKEVKAOLKEAEV 736
 DB 724 HPSEDEKENHAGLNPSADNLKYPSTDTETETEEAEADTTDEAEIPOVENSVINAKIADAE 783
 QY 737 LLAQVTDSSLKANATETLAGLNNLTLOIMDNNSIMAEKLLALLKGSNPSSV 790
 DB 784 LLEKVTDPISIRONAMETLTGLKSLLLGTCKDNNTTISAEDSLALLKESQAPAI 837
 RESULT 26
 ID AAV75934 standard; Protein; 838 AA.
 XX AAV75934;
 AC AAV75934;
 DT 08-MAY-2002 (first entry)
 DE Streptococcus pneumoniae BVH-11-2 protein.
 XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KM pneumonia; streptococcal bacterial infection; BVH-11-2.
 XX Streptococcus pneumoniae.
 OS Streptococcus pneumoniae.
 PN MO200198334-AZ.
 PD 27-DEC-2001.
 PF 19-JUN-2001; 2001WO-CA00908.
 PR 20-JUN-2000; 2000US-212683P.
 PA (SHR-) SHIRE BIOCHEM INC.
 PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX MPI; 2002-122272/16.
 DR N-PSDB; ABL15104.
 XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia
 XX Example 1; Fig 8; 113pp; English.
 PS The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardii or
 CC Streptococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This is the amino acid sequence of
 CC Streptococcus pneumoniae protein BVH-11-2, used to create the antigenic
 CC peptides described in the method of the invention.
 XX Sequence 838 AA;
 SO Query Match 66.0%; Score 2750.5; DB 23; Length 838;
 Best Local Similarity 64.5%; Pred. No. 1.5e-194;
 Matches 538; Conservative 90; Mismatches 145; Indels 61; Gaps 8;

QY 1 SYELGYOARTV-KENNRVSYIDGKOATOKENTLPPEVSKREGINAEQIVIKITDQGV 59
 DB 21 SYELGRHOAGGVKESNRVSTIIDDQAGQKAEENITPPEVSKREGINAEQIVIKITDQGV 80
 QY 60 TSHGDHYHYNGKVPYDAIIEELLKMDPNYKDKEDI VNEVKGYYIKVDGKYVYLKD 119
 DB 81 TSHGDHYHYNGKVPYDAIIEELLKMDPNYKDKEDI VNEVKGYYIKVDGKYVYLKD 140
 QY 120 AAHADNRTKEIRKQOEHSQHREGGTPRDGVALARSQGRITTDGTFNMSDIIED 179
 DB 141 AAHADNRTKEIRKQOEHSQHREGGTPRDGVALARSQGRITTDGTFNMSDIIED 197
 QY 180 TGDAYIVPHGDHYHYIPNELSASELAEEAFISGRGLNSRTFRONSNTSRTWVP 239
 DB 198 TGDAYIVPHGDHYHYIPNELSASELAEEAFISGRGLNSRTFRONSNTSRTWVP 246
 QY 240 SVSNPGTTNTNTSNNSNTNSQASQSDIDSLKQLYKLPSQRHVESDGLVFPDAQITSR 299
 DB 247 NPQVPRLEENHNLVTFTYHQ-NGENISSILRELYAKPLSERHVESDGLVFPDAQITSR 305
 QY 300 TARQVAVPHGDHYHYIPYQMSLEERLARIIPRYSNHWPDPSPRQSPQTPPSPS- 358
 DB 306 TARQVAVPHGDHYHYIPYQMSLEERLARIIPRYSNHWPDPSPRQSPQTPPSPS- 365
 QY 359 -----PGOPAPNLKIDSNSSSLVQLVKYGEVYFPEKGISRYVFAKDIPESETVKULES 413
 DB 366 SLOPAPNPQAPSPID-EKLYKEAVRKQDGVYFEEENGVSRIIPAKDISAETLAAGIDS 423
 QY 414 KLSKQESVSHLTTPAKKENVAPRQDEFYDKAYNLLTEAHKALFNKNGNSDFQALDKLLER 473
 DB 424 KLSKQESVSHLTTPAKKENVAPRQDEFYDKAYNLLTEAHKALFNKNGNSDFQALDKLLER 483
 QY 474 LNDESTNKEKLVLDLLAFAPITPERLGRKNSQIETDEBVRQAOLADYTTSDGYIFD 533
 DB 484 LKDVSSQKVLVDLLAFAPITPERLGRKNSQIETDEBVRQAOLADYTTSDGYIFD 543
 QY 534 EHDIIISDEGAYVYPHNGSHWIKGDSLSDEKRYAAQAYTKEGILPPSPADAVKAPPTG 593
 DB 544 PROITSDGDAYVTPHNTSHWIKGDSLSDEKRYAAQAYTKEGILPPSPADAVKAPPTG 603
 QY 594 DSAAIYINRYKGEKRIPLVRLPYVHEHTVEKGNLIIIPKDHYNIKFAMFDHTYKAP 653
 DB 604 KGAELIYINRYKGEKRIPLVRLPYVHEHTVEKGNLIIIPKDHYNIKFAMFDHTYKAP 663
 QY 654 NGYTLDELPAITIKYVYHPDERPHSNDGNGNASHVUGK-----KDSE----- 697
 DB 664 KGYSLDELPAITIKYVYHPDERPHSNDGNGNASHVUGK-----KDSE----- 723
 QY 698 -----DPKNFKADEEPEVETPAEPEVPOVETKEVKAOLKEAEV 736
 DB 724 HPSEDEKENHAGLNPSADNLKYPSTDTETETEEAEADTTDEAEIPOVENSVINAKIADAE 783
 QY 737 LLAQVTDSSLKANATETLAGLNNLTLOIMDNNSIMAEKLLALLKGSNPSSV 790
 DB 784 LLEKVTDPISIRONAMETLTGLKSLLLGTCKDNNTTISAEDSLALLKESQAPAI 837
 RESULT 27
 ID AAB12759 standard; Protein; 834 AA.
 XX AAB12759;
 AC AAB12759;
 DT 21-NOV-2000 (first entry)
 DE Streptococcus pneumoniae strain RX1 BVH-11-2 protein antigen.
 XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KM pneumonia; streptococcal bacterial infection; BVH-11-2.
 XX Streptococcus pneumoniae.

XX WO200039299-A2.
 PN 06-JUL-2000.
 PD 20-DEC-1999; 99WO-CA01218.
 XX 23-DEC-1998; 98US-0113800.
 PR (BIOC-) BIOCHEM PHARMA INC.
 XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N,
 XX WPI; 2000-452397/39.
 DR Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteremia and/or pneumonia -
 XX Disclosure; Fig 12; 106pp; English.
 XX The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the proteins
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents a
 CC S. pneumoniae BVH-11-2 protein antigen, from the present invention.
 XX
 XX Sequence 834 AA;
 SQ
 Query Match 65.5%; Score 2733; DB 21; Length 834;
 Best Local Similarity 63.5%; Pred. No. 3e-193;
 Matches 538; Conservative 92; Mismatches 145; Indels 72; Gaps 9;
 QY 1 SYELGLYQARTY-KENNRVSYIDGKQATQKENTLPEDEVSKKEGINAEQIVIKITDQGV 59
 DB 2 SYELGHRQAGQVKESNRVSYIDGDAQKENTLPEDEVSKKEGINAEQIVIKITDQGV 61
 QY 60 TSHGDHYHYNGKVPYDAIISEELMKDPNYKDKEDIYNEVGKGYIVKDGKYYVYKLD 119
 DB 62 TSHGDHYHYNGKVPYDAIISEELMKDPNYKDKEDIYNEVGKGYIVKDGKYYVYKLD 121
 QY 120 AAHADNVRTKEELNRKQESQHREGTFRNGAVALLASQGRYTTDDGYTFNASTIID 179
 DB 122 AAHADNVRTKEELNRKQESQHREGTFRNGAVALLASQGRYTTDDGYTFNASTIID 178
 QY 180 TGDVAVIPHGDHYHYIPKNELSASELAAEAFLSGRGNLSRTYRQNSDTSRTNWPV 239
 DB 179 TGDVAVIPHGDHYHYIPKNELSASELAAEAFLSGRGNLSRTYRQNSDTSRTNWPV 227
 QY 240 SVSNPGTNTNTSNNSNTNSQASQSDIDSLKQLYKPLPSORHVESDGLVDPAGITSR 299
 DB 228 NPAQPLSLSENHNLVTPYHQ-NOGENISLRELTAKEPLSRHVESDGLVDPAGITSR 286
 QY 300 TARGVAVPHGDHYHYIPYSGOMSELEERLARIITPLRRSNHWVPDSRPEQSPQTPPEBP 359
 DB 287 TAMGVAVPHGDHYHYIPYSGOMSELEERLARIITPLRRSNHWVPDSRPEQSPQTPPEBP 346
 QY 360 GPQAPNLK-IDSN--SSLVQVLRKVGEGYFEESKISRIVFAFDLSEFVKNLESL 415
 DB 347 SQQAPNPQAPSNPIDKLVKEAVRKVDGYFEENGPRITIPADLAEFTAGIDSKL 406
 QY 416 STQESVSHTLTAKKENAVRDOEFYDKAYNLLTEAHKLFXXKGRNSDQALDKLERLN 475
 DB 407 AAOESLSHKLGAKKTDLPSSDREFYKAYDLARIHODLDMNGRQVDEALDNLERL 466
 QY 476 DESTKKEKYVDLLAFIARTIPERLGRNSQIEYFEDSVRAQLADKTTSDGYTFD 535
 DB 467 DVSSDVKLVVDILAFIRPERLGRNSQIEYFEDSVRAQLADKTTSDGYTFD 526
 QY 536 DIISDEGDAYVTPHMGSHWIGKDSLDEKEXVAAQAYTEKGLIPSPADVYKANPTGDS 595

DB 527 DITSEGDYAVTPHMGSHWIGKDSLDEKEXVAAQAYTEKGLIPSPADVYKANPTGDS 586
 QY 596 AAATNRPKGEKRIILVLPYVTEHTVEYKGNLLIIPHDHYNHINKFAMFDHTTKAPNG 655
 DB 587 AAATNRPKGEKRIILVLPYVTEHTVEYKGNLLIIPHDHYNHINKFAMFDHTTKAPNG 646
 QY 656 YTEDELPAITIKYVHPDERPHSNDGMSYSHVIGKK-----DHSEDPNK----- 701
 DB 647 YTEDELPAITIKYVHPDERPHSNDGMSYSHVIGKK-----DHSEDPNK----- 700
 QY 702 -----NFKAD-----EPPVETPAEPVPOVE 723
 DB 707 EEDKEHDEVSEPTHPDESDEKENHVGILNPSADNLVYKPTDTEETEEADDTDEAIPQVE 766
 QY 724 TKVNAQLKEAEVLLAKTDSLSKANATETLGLRNNTLQIMDNNSIMAEKELALIK 783
 DB 767 YSVINAKIAEAEALLKXTDSSIRONAVETTLGLSGTGDNNNTISAEVDSLALIK 826
 QY 784 GSNPSSV 790
 DB 827 ESQAPPI 833
 RESULT 28
 AAB12756
 ID AAB12756 standard; Protein; 816 AA.
 AC AAB12756;
 XX 21-NOV-2000 (first entry)
 XX
 DE Streptococcus pneumoniae strain P4241 BVH-11-2 protein antigen.
 XX
 KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
 KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
 KW otitis media; pneumonia; immunisation; bactericidal.
 OS Streptococcus pneumoniae.
 XX
 PN WO200039299-A2.
 XX 06-JUL-2000.
 PD 20-DEC-1999; 99WO-CA01218.
 PR 23-DEC-1998; 98US-0113800.
 XX (BIOC-) BIOCHEM PHARMA INC.
 PA Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N,
 PI WPI; 2000-452397/39.
 DR Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteremia and/or pneumonia -
 XX Disclosure; Fig 12; 106pp; English.
 XX The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from Streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of Streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents a
 CC S. pneumoniae BVH-11-2 protein antigen, from the present invention.
 XX
 XX Sequence 816 AA;
 SQ
 Query Match 65.5%; Score 2728; DB 21; Length 816;
 Best Local Similarity 63.9%; Pred. No. 6.7e-193;

Matches	530:	Conservative	100:	Mismatches	144:	Indels	56:	Gaps	7									
QY	1	SYELGLYQA-R	TVKENNRVSYIDGQAO	TKENTLPDEVSKREG	INAEQIVIKITDQGY	59												
Db	2	SYELGRHQAGQD	KKESNRYVAVIDGQAG	CAKENTLPDEVSKREG	INAEQIVIKITDQGY	61												
QY	60	TSHGHHYHYNK	VPYDAIISEELLMK	DPNNYTKDEDI	ENYKGGVIVVDKRYVYLKD	119												
Db	62	TSHGHHYHYNK	VPYDAIISEELLMK	DPNNYTKDEDI	ENYKGGVIVVDKRYVYLKD	121												
QY	120	AAHADNVR	TKEEINRQKOE	SHOSHRREGGTR	PNDAVALARSQGRYTTDDGY	179												
Db	122	AAHADNIR	TKKEIKQKQEH	SHNHGGS--NQAVAA	AQGRYTTDDGY	179												
QY	180	TGDAYIVPHG	DHYHIIPKNELS	ASBELAAEAFLSG	RGNLSNRTYRQNSDNT	239												
Db	180	TGDAYIVPHG	NHFHYIPKSDLS	ASBELAAEAQVWNG--	-----KQGSRESSSSHNA	228												
QY	240	SVSNGTNTNT	NSNNTNSQASQ	SNDDISLLKOLYKPL	SLQRHVSDCLVDP	299												
Db	229	NPAQR	LSNNHLVTYPT	YHQ--NOGENIS	LSRELAKPLSRHVESDGL	179												
QY	300	TARGA	VPFHGDYHFI	PYSQMS	ELERARIIPLRKSNHW	VPDSRPEQSPQ--PTPEP	357											
Db	288	TARGA	VAEPHGNHYHFI	PYQMS	ELERARIIPLRKSNHW	VPDSRPEQSPQSP	347											
QY	358	SPGP	PAPRLKID	SNSSLV	QLVRKVEG	VEEKGISRYVP	AKDLPSETV	NKLSKLSK	417									
Db	348	APNPQ	PAPENPID--	EKLVEA	VRKVGDCG	YVEENVSRYIIPAKDLS	ABTA	GIDSKLAK	405									
QY	418	QESV	HTLTAKKEN	VA	PRDOE	FYDKAYNNLT	EAHKLFPKNG	RNSPQALDKLER	477									
Db	406	QESL	HKLGTKT	TDLPSSD	REFYNNKAYD	LARIHODL	LNKRQYD	FEALDNL	LEBLKV	465								
QY	478	STNKE	LVDDLAFL	APITHP	ERLKG	PNSQIE	XTED	EVIAI	QADYTS	537								
Db	466	SSDK	KLVEDLAFL	APINHP	ERLKG	PNSQI	YTTD	EIOVAK	LAGK	YTTEDGY	525							
QY	538	ISDEG	DAVYTPH	MGSHWIG	KDSL	SDKE	KVAQA	AYTKE	KGILPR	PSB	DAVKANPT	597						
Db	526	TSDEG	DAVYTPH	MTSHWIK	KD	SLSE	AEBA	AAQA	VAKE	GLPR	STDH	DSGNT	585					
QY	588	AIYN	RVKGEK	RIPLVRL	PMVE	TYE	VKGNLIIP	PKDH	YANNIK	RAM	PD	HTYK	657					
Db	586	AIYN	RVKAKK	APL	RLMP	PNLI	QY	TE	VEK	NGSLIIP	YD	HYNI	645					
QY	658	LED	L	FAT	II	YU	YEH	PER	PHS	NDG	MNA	SEH	693					
Db	646	LED	L	LA	T	AT	YU	YEH	PER	PHS	NDG	MNA	705					
QY	694	-----	DHSE	DPN	KFK	PD	DEE	VEE	TE	PA	PE	VO	740					
Db	706	DEK	EN	HA	GL	PN	SAD	NLY	K	PS	D	T	765					
QY	741	VTD	S	L	K	A	N	A	T	T	L	A	790					
Db	766	VTD	P	S	I	R	O	N	A	T	T	L	815					
RESULT 29																		
AAB12757																		
ID AAB12757																		
AAAB12757;																		
AC AAB12757;																		
AAAB12757;																		
DT 21-NOV-2000																		
XX (first entry)																		

OS Streptococcus pneumoniae.
XX
XX
PN WO200039299-A2.
XX
XX 06-JUL-2000.
XX
XX 20-DEC-1999; 99WO-CA01218.
XX
XX 23-DEC-1998; 98US-0113800.
XX
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
XX
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteremia and/or pneumonia -
XX
XX
PS Disclosure; Fig 12; 106pp; English.
XX
XX The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the proteins
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteremia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BvH-11-2 protein antigen, from the present invention.
XX
XX Sequence 816 AA;
SQ

Query Match	65.5%	Score 2727	DB 212	Length 816
Best Local Similarity	63.7%	Pred. 7.9e-193		
Matches	529	Conservative 101	Mismatches 144	Indels 56
				Gaps 7
Qy	1 SYELGLYOA-RIVKNNRNSVYIDGKQAOIKTENTLPDEVSKREGINAEDIVIKITDQGVY	59		
Db	2 SYELRRHQAGQDKKESNRVAIYIDGQAGKAKENLTPDEVSKREGINAEDIVIKITDQGVY	61		
Qy	60 TSHGDHYHYNGKQVYDAIISEELLMKDPNYLKDIEDIVNEVKGQVIVKDGKYYVYLLKD	119		
Db	62 TSHGDHYHYNGKQVYDAIISEELLMKDPNYLKDIEDIVNEVKGQVIVKDGKYYVYLLKD	121		
Qy	120 AAHADNVRTKEEINFQKQESHQSHREGGIPRNDGAVALLASQGRYTTDDGIFINASDIIED	179		
Db	122 AAHADNIRTKKEIKKROEHSNHNGGGS--NDQAVVAARAQGRYTTDDGIFINASDIIED	179		
Qy	180 TGDATVPHGSHYHYIPIKNSLSASLSAAEAFLSGRGNLSNRTYRQKSDMSTRNWAP	239		
Db	180 TGDATVPHGSHYHYIPIKSDLSASLSAAQAQWNG-----KQGRSPSSSSSHNA	228		
Qy	240 SVSNPGTNTNTNNSNNTNSQASNDIDSLIKQLYKLPLSGRHESDGLVDFDPAQITSR	299		
Db	229 NPAQRIEENHNLVITPTTHQ--NGEENISLLRELYAPLSRHHVSDGLIDFPAQITSR	287		
Qy	300 TARGAAYVPHGDHYHPIPYSSOMSELEERLARIIPLRYSRNHWVPDSRPEQSPSQ--PTPEP	357		
Db	288 TARGAAYVPHGSHYHYIPIEQMSELEERLARIIPLRYSRNHWVPDSRPEQSPSPSQP	347		
Qy	358 SPGPQAPAPNLKIDSNSLSVSLVRKVGEGYVFEKGISRYVFAKDLPSSTVKNLSKLSK	417		
Db	348 APNPQAPASNPID--EKLVKERAVRKVGEGYVEENGVSRYIPAKDLASFTAGIDSKLAK	405		
Qy	418 QESVSHITLAKKENAVPRDQEFYDAYNULTEAHKALPFRKNRNDFOALDKLLESLNE	477		
Db	406 QESLSHKIKGTKTDTLPSDDREFYNAAYDULARIHDDLNNKRQYDFEALDNLLESLKQV	465		
Qy	478 STNKSGLVDDLAFIAPITHPERLCKPNSQLEYTEDEVRIAOLADKYTSSDGYIPDEHDI	537		
Db	466 SSDKKLVEDIILAFIAPIRHPERLCKPNSQIYITDDEIOVAKLAKGKITTEDGYIPDPRI	525		

QY 538 ISDEGDAVYTPHMGSHWIGKDSLSDKEKVAQAAYTKENGILPPSPDADVKANPTGDSAA 597
DB 526 TSDGDAVYTPHMTTHSHWIKKDSLSEABRAAAQAYAKKEGILTPSPDHDQSGNTEAKGAE 585
QY 598 AIYNRYKGEKRIPLVRLPYVWEHTVEVKGNLIIIPKDHYNHIKFAFPDHTYKAPNGYT 657
DB 586 AIYNRYKAAKVPPLDMPPYNLQYTVVEKNGSLIIIPHYDHYNIKFEPFDEGLYEAPKGYT 645
QY 658 LEDLFATIKYVYEHDPERPHSNDGWNASEHVLGKK----- 693
DB 646 LEDLATVYKYVYEHDPERPHSDNGFGNASHVAKKADQDSKPEDKGHDEVSEPTHPES 705
QY 694 -----DHSEDPKNFKADEBPVEETPAPEVPEQVETEKEVAQLKEAEVILAK 740
DB 706 DEKENHAGLNPSADNLYKPSDTETETEEBAEDTTDEAEIIPQVESHVINAKIADAEALLEK 765
QY 741 VTDSSIKANATETLAGLRNNLTLOIMDNNSIMAEAEKILALKGSNPSV 790
DB 766 VTDPSIRQNAEMETLGLKSSLLGTGKDNNTISAEVDSLALLKKSQAPAI 815
RESULT 30
AAB12760
ID AAB12760 standard; Protein; 811 AA.
XX
AC AAB12760;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae strain P4241 BVH-11 protein antigen.
XX
KM Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.
XX
PD 06-JUL-2000.
XX
PF 20-DEC-1999; 99MO-CA01218.
XX
PR 23-DEC-1998; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
DR WPI; 2000-452397/39.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX
PS otitis media, bacteraemia and/or pneumonia -
XX
XX Disclosure; Fig 12; 106pp; English.
CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BVH-11 protein antigen, from the present invention.
XX
SQ Sequence 811 AA;
Query Match 65.3%; Score 2720; DB 21; Length 811;
Best Local Similarity 64.1%; Pired. No. 2.6e-192;
Matches 529; Conservative 98; Mismatches 142; Indels 56; Gaps 7;
QY 1 SYELGUYQA-RTVKENNRVSYIDGKQATQKTENLTPEVSKREGINAEQIVIKITDQGVY 59

DB 2 SYELGHRHQGDQKSSNRVAYIDGDAQKAENLTPEDEVSKREGINAEQIVIKITDQGVY 61
QY 60 TSHGDHYHYNNKVPDPAIISELLMKDPNYKLDKEDIYNEVKGVIVKVDGYYVYLD 119
DB 62 TSHGDHYHYNNKVPDPAIISELLMKDPNYKLDKEDIYNEVKGVIVKVDGYYVYLD 121
QY 120 AAHADNVRKEEINRQKQSHQREGGTPRNGAVALARSOGRYTTDDGYIFNASDIID 179
DB 122 AAHADNIRKKEELIKQKQKQSHHNGGGS--NDAVAVARAQGRYTTDDGYIFNASDIID 179
QY 180 TGDAYIVPHGDHYHYI PKNELSASELAAAEPLSGRNLNSRTRYRQNSDNTSRTNWP 239
DB 180 TGDAYIVPHGNPHHYI PKSDLDASELAAQAAYWNG-----KQGRSSSSSHNA 228
QY 240 SVSNPGTTNTNTSNNSNTNSQASQSDIDSLKOLYKPLSGRPHYSDGLVDPQAOTSR 299
DB 229 NPAQPLSENHNLTYTPYHQ--NOGENTSSLRRELYAKPLSRHYESDGLIFDPAQITSR 287
QY 300 TARGVAVPHGDHYHFPYSQMSLEERLARIIPLRYSNHWYVDPSPQSPQ--PTPEP 357
DB 288 TARGVAVPHGNHYHFIYEQMSELEBRILARIIPLRYSNHWYVDPSPQSPQSPQ 347
QY 358 SPGPQAPNLKIDNSSILVSQLVRKVGEGYVEEKGISRYVPAKDLPSRTVKNLSKLSK 417
DB 348 APNPQAPENPID--EKLVEAVRKVGCGDYVEENGVSRYIPAKOLSAETAGIDSKLAK 405
QY 418 QESVSHTLTAKKENVAAPRQEFYDKAYNLTLTAHKALEFNKGRNSDQALDKLERLND 477
DB 406 QESLSHKLGTKKTDLPPSSDREFFYNKAYDLARIHODLLDNKGRQVDFEALDNLLERLKDV 465
QY 478 STNKEKLVNDLAFAPIRIHPERLGNQIYETEVEVIAQADKYTSDGYTFPEHDI 537
DB 466 SSKKALVEDIILAFAPIRIHPERLGNQIYETEVEVIAQADKYTSDGYTFPEHDI 525
QY 538 ISDEGDAVYTPHMGSHWIGKDSLSDKEKVAQAAYTKENGILPPSPDADVKANPTGDSAA 597
DB 526 TSDGDAVYTPHMTTHSHWIKKDSLSEABRAAAQAYAKKEGILTPSPDHDQSGNTEAKGAE 585
QY 598 AIYNRYKGEKRIPLVRLPYVWEHTVEVKGNLIIIPKDHYNHIKFAFPDHTYKAPNGYT 657
DB 586 AIYNRYKAAKVPPLDMPPYNLQYTVVEKNGSLIIIPHYDHYNIKFEPFDEGLYEAPKGYT 645
QY 658 LEDLFATIKYVYEHDPERPHSNDGWNASEHVLGKK----- 693
DB 646 LEDLATVYKYVYEHDPERPHSDNGFGNASHVAKKADQDSKPEDKGHDEVSEPTHPES 705
QY 694 -----DHSEDPKNFKADEBPVEETPAPEVPEQVETEKEVAQLKEAEVILAK 740
DB 706 DEKENHAGLNPSADNLYKPSDTETETEEBAEDTTDEAEIIPQVESHVINAKIADAEALLEK 765
QY 741 VTDSSIKANATETLAGLRNNLTLOIMDNNSIMAEAEKILALKGS 785
DB 766 VTDPSIRQNAEMETLGLKSSLLGTGKDNNTISAEVDSLALLKES 810
RESULT 31
AAB12762
ID AAB12762 standard; Protein; 811 AA.
XX
AC AAB12762;
XX
DT 21-NOV-2000 (first entry)
XX
DE Streptococcus pneumoniae strain A66 BVH-11 protein antigen.
XX
KM Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX
XX
OS Streptococcus pneumoniae.
XX
PN WO200039299-A2.

```

XX 06-JUL-2000.
PD 20-DEC-1999; 99WO-CA01218.
XX 23-DEC-1998; 98US-0113800.
XX (BIOC-) BIOCHEM PHARMA INC.
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteremia and/or pneumonia -
XX Disclosure; Fig 12; 106pp; English.
XX The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens
XX have bactericidal activity. The nucleic acids, encoding the proteins
XX antigens, may be used for the recombinant production of the proteins
XX they encode. The protein antigens may then be used as vaccines for the
XX prevention and treatment of Streptococcal infections in mammals
XX (especially humans) which result in, e.g. meningitis, otitis media,
XX bacteraemia and/or pneumonia. The present sequence represents a
XX S. pneumoniae BVH-11 protein antigen, from the present invention.
SQ Sequence 811 AA;
Query Match 65.3%; Score 2720; DB 21; Length 811;
Best Local Similarity 64.1%; Pred. No. 2,6e-192;
Matches 529; Conservative 98; Mismatches 142; Indels 56; Gaps 7;
OY 1 SYELGLYQA-RTVKNNRVSYIDGKQATQKTNLTPEVSKREGINAEQIVIKITDQGYV 59
DB 2 SYELGRHOAGODKESNRVAYIDGDAQCKAENLTPEVSKREGINAEQIVIKITDQGYV 61
OY TSHGHYHYNGKVPYDAIISEELMKDPNYKLKXEDIVNEVKGKGYVKKVGKYYVYLKD 119
DB 62 TSHGHYHYNGKVPYDAIISEELMKDPNYQLKOSDIYNEIKGGYVKKVGKYYVYLKD 121
OY 120 AAHADNVRTEKIEINQKQESHQHREGTTPRNDGAVALARSOGRYTTDDGYIFNADITD 179
DB 122 AAHADNITTEKIEIKQKQESHNNHGGG--NDQAVVAAQAGRYTTDDGYIFNADITD 179
OY 180 TGDATYVPHGHYHYIPKXELASASLAAEAFLSGRNLNSRTYRRONSNTSRTNWP 239
DB 180 TGDATYVPHGHYHYIPKSDISASLAAQAAYWNG-----KQSGRPSSSSSHNA 228
OY 240 SVSNGTNTNTSNNNSQASNDIDSLKOLYKPLSGRHVESGLVDPDPOITSR 299
DB 229 NPAQRLESENHNLVTPYHQ--NOGENISSLRELYAKPLSRHVESGLVDPDPOITSR 287
OY 300 TARGAVPHGHYHYIPYSOMSELEERARIILPLRYRSNNHWPPDSRPQSPQ--PTREP 357
DB 288 TARGAVPHGHYHYIPYEQNSELEERARIILPLRYRSNNHWPPDSRPQSPQSPQ 347
OY 358 SPGPAPNALKIDSNSLSVQLVRKVGEGYVEEKGIRRYFAKLPSETVKNLESKL 417
DB 348 APNPAPASNPID--EKLVEKAVRKVGEGYVEEKGIRRYIPAKLSAETAGIDSKLAK 405
OY 418 QESVSHTLTAKKENVAPRDOEFYDCAVNLTEAHKALPXNKGKNSDPALDLERL 477
DB 406 QBSUSHKIGTKTDTDPSSDRFYNKAYIDLARIHODLLDNKRQVDFELDLERLKV 465
OY 478 STNKEKLVDDLAFAPITPHERLQKPNQSELYTDEVARIAQLADKYTTDGYIDEHDI 537
DB 466 SSDKVLVEDILAFAPIRHPERLQKPNQSELYTDEVARIAQLADKYTTDGYIDEHDI 525
OY 538 ISDEGDVYTPHMGSHWIGKDSLDEKQVAAQAYTKEGILPSPDADVYKAPPTGSDAA 597
DB 526 TSDEGDVYTPHMTSHWIKKDSLSEARAAQAAYAKEKGLTPPSTDHDSGNTAKGAE 585

```

```

OY 598 AIYRVKGEKRIPLVRLPYMEHTVEVANGNLIPHKDHYNIKFAMFDDHTYKAPNGYT 657
DB 586 AIYRVKAAKRVPLDRFPYNIQYVEVANGSLIIPHDHYNIFEFMFDEBLYEAPKGYT 645
OY 658 LEDLPAITKYVVEHPDERPHSNDGWNASEHVLGKK----- 693
DB 646 LEDLPAITKYVVEHPDERPHSNDGWNASEHVLGKK----- 693
OY 694 -----PHSEDPNNKPKADEPVEPTPAPEPQVETEKVEAOJKEAEVLLAK 740
DB 706 DEKENHAGLNSADNLKVPSTDTETEBEADYTDDEABIPVEHVSINAKTADABALUEK 765
OY 741 VTDSILKANATETTLAGLRNNLTQIMDINSIMAEAKLLALLKGS 785
DB 766 VTDSIRQNAEMETILGLKSSLLGTXKNNNTISAEVDSLALLKES 810
RESULT 32
AAB12758
ID AAB12758 standard; Protein; 816 AA.
AC AAB12758;
XX 21-NOV-2000 (first entry)
XX Streptococcus pneumoniae strain WU2 BVH-11-2 protein antigen.
XX Streptococcus pneumoniae, BVH-3; BVH-11, BVH-28; antigen; vaccine;
XX prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX otitis media; pneumonia; immunisation; bactericidal.
XX Streptococcus pneumoniae.
XX WO200039299-A2.
XX 06-JUL-2000.
XX 20-DEC-1999; 99WO-CA01218.
XX 23-DEC-1998; 98US-0113800.
XX (BIOC-) BIOCHEM PHARMA INC.
XX Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX WPI; 2000-452397/39.
XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
XX otitis media, bacteremia and/or pneumonia -
XX Disclosure; Fig 12; 106pp; English.
XX The present invention describes nucleic acids (I) encoding protein
XX antigens (II) from Streptococcus pneumoniae. The protein antigens
XX have bactericidal activity. The nucleic acids, encoding the proteins
XX antigens, may be used for the recombinant production of the proteins
XX they encode. The protein antigens may then be used as vaccines for the
XX prevention and treatment of Streptococcal infections in mammals
XX (especially humans) which result in, e.g. meningitis, otitis media,
XX bacteraemia and/or pneumonia. The present sequence represents a
XX S. pneumoniae BVH-11-2 protein antigen, from the present invention.
XX Sequence 816 AA;
Query Match 65.2%; Score 2714; DB 21; Length 816;
Best Local Similarity 63.6%; Pred. No. 7.3e-192;
Matches 528; Conservative 100; Mismatches 146; Indels 56; Gaps 7;
OY 1 SYELGLYQA-RTVKNNRVSYIDGKQATQKTNLTPEVSKREGINAEQIVIKITDQGYV 59
DB 2 SYELGRHOAGODKESNRVAYIDGDAQKAEENLTPEVSKREGINAEQIVIKITDQGYV 61

```

```

QY 60 TSHGDHYHYNGKVPYDAIISELLMKDPNYKLDKEDIVNEVKGQYVIRKVGKYVYLLKD 119
XX |||||
DB 62 TSHGDHYHYNGKVPYDAIISELLMKDPNYKLDKEDIVNEVKGQYVIRKVGKYVYLLKD 121
QY 120 AAHADNVRTKEEINRQKQESHQREBGTFRNDGAVALARSGQRYTTDDGYIFNASTDIID 179
XX |||||
DB 122 AAHADNVRTKEEIKRQKQESHQSHHGGS--NDQAVVAARAQGYTTDDGYIFNASTDIID 179
QY 180 TGDYAVPFGHDHYHFKPKEELASAEFLASAEFLSGNLSNRTYRQNSDNTSRNWP 239
XX |||||
DB 180 TGDYAVPFGDNHFIKPSDLSASELAQAQWNG-----KQSGRSSSSSHNA 228
QY 240 SVSNPGTTNTNNTSNNTNSQASQSDIDSLKQLYKPLSQRHVESDGLVDPQAQITSR 299
XX |||||
DB 229 NPAQPRLSNHLVTYPTTHQ--NQGEMISSLARELVAKPLSERHVESDGLVDPQAQITSR 287
QY 300 TARGAVAVPHGDHYHFIPIYSQMSLEBRIARIIPLRYSNHWVPDSRPEQSPQ--PTPEP 357
DB 288 TARGAVAVPHGNHYHFIPIYQMSLEBRIARIIPLRYSNHWVPDSRPEQSPQSPQ 347
QY 358 SPGPQAPRLKIDNSLSVQLVRKYGEGVEEKGISRYVFAKDLPSFTVKNLESKLSK 417
DB 348 ANPQAPRPNPID--EKLVEARAKVGQGVVEENGVSRYIPAKDLASATAAGIDSKLAK 405
QY 418 QESVSHTLTAKKENVAPRDQEFYDKAYNLLTEAHKALFKNGKNSDFQALDKLELND 477
DB 406 QESLSHKLQTKTDLPSRREFYNKAYDLARIHOLLNKRQVDFEALDNLLEKLV 465
QY 478 STNKEKLVDDLAFIAPITHPERLQKPNQIEYTEDEVRIAQIADKYTTSDGYIFDEHDI 537
DB 466 SSDKVLVEDILAFIAPIRHPERLQKPNQIYTTDEIQVAKLAGKYTTEDGYIFDPRI 525
QY 538 ISDEGDAYVTPHMGSHWIGKDSLDEKEVAQAQYTKEGILPSPDAVKNPTDSQA 597
DB 526 TSDGDAYVTPHMTSHHWIKKDSLSEAEPAQAQYKEGILTPSTDHDSGNTAKGAE 585
QY 598 AIYNRYKGEKRIPLVRLPYMVEHTVEVKNGNLIIPKADHYHNIFKAMPDHTYKANGYT 657
DB 586 AIYNRYKAKKVPFLDMPYNLQYTVENVKGSLLIIPHYDHYHNIFKEMFEGLYEAKGYT 645
QY 658 LEDLFATIIYVYEHDPERPHSDNGWGNASEHVLGKK----- 693
DB 646 LEDLATATVYVYEHDPERPHSDNGWGNASDHVKNKADQSKPDEKGDHDEVESEPTHPES 705
QY 694 -----DHSEDPNKNFKADEEPEVETPAPEVPQVTEKVEAQLEAVLAK 740
DB 706 DEKENHAGINPSADNLKPKSTDEETEEDAEEDTTDEAEIPOVEHSVINKIADAEALLEK 765
QY 741 VTDSSLKANATETIAGLRNNLTQIMDNNSIMAEKLLALLKGSPPSSV 790
DB 766 VTDPSIRQNAETLTGLKSSLLGTGKDNNTISAEDVSLALLKKSQAPAI 815

```

RESULT 33
AAB12761
ID AAB12761 standard; Protein; 811 AA.

```

XX AAB12761;
XX
XX 21-NOV-2000 (first entry)
XX
XX Streptococcus pneumoniae strain WU2 BVH-11 protein antigen.
XX
XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
XX prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
XX otitis media; pneumonia; immunisation; bactericidal.
XX
XX Streptococcus pneumoniae.
XX
XX WO2000039299-A2.
XX
XX 06-JUL-2000.
XX

```

```

PF 20-DEC-1999; 99WO-CA01218.
XX
PR 23-DEC-1999; 98US-0113800.
XX
PA (BIOC-) BIOCHEM PHARMA INC.
XX
PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
XX
DR WPI; 2000-452397/39.
XX
PT Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia -
XX
PS Disclosure; Fig 12; 106pp; English.
XX
CC The present invention describes nucleic acids (1) encoding protein
CC antigens (11) from Streptococcus pneumoniae. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of Streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents a
CC S. pneumoniae BVH-11 protein antigen, from the present invention.
XX
SQ Sequence 811 AA.

```

Query Match 65.1%; Score 2713; DB 21; Length 811;
Best Local Similarity 64.0%; Pred. No. 8,6e-192;
Matches 528; Conservative 98; Mismatches 143; Indels 56; Gaps 7;

```

QY 1 SYELGLYQA-RIVKNNRVSYYDGOATQKTENLTPDEVSKREGINAEQIVKITDQGV 59
DB 2 STELGRHQGGQDKSSNRVAVIDGDAQKAENLTPDESKREGINAEQIVKITDQGV 61
QY 60 TSHGDHYHYNGKVPYDAIISELLMKDPNYKLDKEDIVNEVKGQYVIRKVGKYVYLLKD 119
DB 62 TSHGDHYHYNGKVPYDAIISELLMKDPNYKLDKEDIVNEVKGQYVIRKVGKYVYLLKD 121
QY 120 AAHADNVRTKEEINRQKQESHQREBGTFRNDGAVALARSGQRYTTDDGYIFNASTDIID 179
DB 122 AAHADNVRTKEEIKRQKQESHQSHHGGS--NDQAVVAARAQGYTTDDGYIFNASTDIID 179
QY 180 TGDYAVPFGHDHYHFKPKEELASAEFLASAEFLSGNLSNRTYRQNSDNTSRNWP 239
DB 180 TGDYAVPFGDNHFIKPSDLSASELAQAQWNG-----KQSGRSSSSSHNA 228
QY 240 SVSNPGTTNTNNTSNNTNSQASQSDIDSLKQLYKPLSQRHVESDGLVDPQAQITSR 299
DB 229 NPAQPRLSNHLVTYPTTHQ--NQGEMISSLARELVAKPLSERHVESDGLVDPQAQITSR 287
QY 300 TARGAVAVPHGDHYHFIPIYSQMSLEBRIARIIPLRYSNHWVPDSRPEQSPQ--PTPEP 357
DB 288 TARGAVAVPHGNHYHFIPIYQMSLEBRIARIIPLRYSNHWVPDSRPEQSPQSPQ 347
QY 358 SPGPQAPRLKIDNSLSVQLVRKYGEGVEEKGISRYVFAKDLPSFTVKNLESKLSK 417
DB 348 ANPQAPRPNPID--EKLVEARAKVGQGVVEENGVSRYIPAKDLASATAAGIDSKLAK 405
QY 418 QESVSHTLTAKKENVAPRDQEFYDKAYNLLTEAHKALFKNGKNSDFQALDKLELND 477
DB 406 QESLSHKLQTKTDLPSRREFYNKAYDLARIHOLLNKRQVDFEALDNLLEKLV 465
QY 478 STNKEKLVDDLAFIAPITHPERLQKPNQIEYTEDEVRIAQIADKYTTSDGYIFDEHDI 537
DB 466 SSDKVLVEDILAFIAPIRHPERLQKPNQIYTTDEIQVAKLAGKYTTEDGYIFDPRI 525
QY 538 ISDEGDAYVTPHMGSHWIGKDSLDEKEVAQAQYTKEGILPSPDAVKNPTDSQA 597
DB 526 TSDGDAYVTPHMTSHHWIKKDSLSEAEPAQAQYKEGILTPSTDHDSGNTAKGAE 585
QY 598 AIYNRYKGEKRIPLVRLPYMVEHTVEVKNGNLIIPKADHYHNIFKAMPDHTYKANGYT 657

```

```
Db      586 AAYNEVKAQKAVPLDRMPYNIQYVEVKNGSLIIPHYDHNHNIKEFWEDEGLYEAPKGYT 645
QY      658 LEDLFATIKYVYEHDPDRPHSNDGNGNASEHVLGKK----- 693
Db      646 LEDLATATYKYVYEHDPDRPHSNDGNGNADHVRKKNKADODSKPDEDKGHDVESEPTHEBS 705
QY      694 -----DHSEDPKNKFKADEPVEETPAPEVPOVETEKVEAOLKEAEVLAK 740
Db      706 DEKENHAGLNPASDNLKRPSTDETEEBEAEEDTTDEAEIPQVEHGVINAKIKADAEALIEK 765
QY      741 VTDSSLKANATETTLAQRNNLTLOIMDNNSIMAEAEKLLALLKGS 785
Db      766 VTDPSIRQVAMETTLGLKSSLLTGKXNNNTTISAEVDSLALLKES 810

RESULT 34
ID      AAM55095 standard; Protein; 763 AA.
XX      AAM55095;
AC      AAM55095;
XX      02-OCT-1998 (first entry)
DE      Streptococcus pneumoniae SP0042 protein.
XX      Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW      detection; pneumonia; otitis media; meningitis.
XX      Streptococcus pneumoniae.
OS      Streptococcus pneumoniae.
XX      Key Location/Qualifiers
FH      Misc-difference 51 /label= unknown
FT      /note= "encoded by ATN"
FT      Misc-difference 469 /label= unknown
FT      /note= "encoded by NCA"
FT      Misc-difference 477 /label= unknown
FT      /note= "encoded by GAN"
XX      WO9818930-A2.
PN      07-MAY-1998.
XX      30-OCT-1997; 97WO-US19422.
PF      31-OCT-1996; 96US-0029960.
PR      (HUMA-) HUMAN GENOME SCI INC.
PA      (HUMA-) HUMAN GENOME SCI INC.
PI      Choi GH, Hromockyj A, Johnson LS, Kunsch CA;
XX      MPI; 1998-272224/24.
DR      N-PSDB; AAV27356.
XX      Nucleic acid encoding antigenic peptide(s) from Streptococcus
PT      pneumoniae - or their epitope-containing fragments, useful in
PT      protective or therapeutic vaccines, and for diagnosis
XX      Claim 11; Page 62; 118pp; English.
XX      The present sequence represents a protein from Streptococcus pneumoniae.
XX      The nucleic acid sequence encoding the Streptococcus pneumoniae protein
XX      can be useful in vaccines for inducing protective antibodies against
XX      Streptococcus pneumoniae, for treatment or prevention of infection e.g.
XX      pneumonia, otitis media or meningitis. Probes based on the nucleic acid
XX      are used to detect Streptococcus infection (by usual hybridisation or
XX      amplification methods), also for isolating Streptococcus genes or their
XX      allelic variants. The protein can be used similarly for detecting specific
XX      antibodies in standard immunoassays, especially for diagnosing or
XX      monitoring infections. Antibodies which bind the protein are used to
XX      detect corresponding antigens, to purify the protein and for passive
```

```
CC      immunisation (optionally coupled to a toxin). Vaccines are administered,
CC      e.g. by injection, orally or through the skin, typically at 0.01-1000
CC      (especially 10-300) mu g/ml per dose.
SQ      Sequence 763 AA;
Query Match 63.6%; Score 2649.5; DB 19; Length 763;
Best Local Similarity 66.8%; Pred. No. 3.9e-187;
Matches 519; Conservative 77; Mismatches 132; Indels 49; Gaps 8;
QY      1 SYEIGLVQARIV-KENNRVSIIDGKQATQKENTLPBEVSRREGINAEQVIYIKITDQGY 59
Db      2 SYEIGRHQAGGVKESNRVSYIDDOAQKQKENTLPBEVSRREGINAEQVIYIKITDQGY 61
QY      60 TSHGDHYHYNGKVPYDAIIEELLMKDPNYKLDEPDIYNEVKGVIKVDGKYVYVKD 119
Db      62 TSHGDHYHYNGKVPYDAIIEELLMKDPNYKLOSDIYNEIKGGYIKVKGKYYVYVKD 121
QY      120 AAHADNVTKKEINROKQESHQREGGTPRNDGAVALARSGQRYTTDGYIFNADIIED 179
Db      122 AAHADNIRTKKEIKROKQESHQREHNS--RADNVAARAGQRTTDDGYIFNADIIED 178
QY      180 TGDAYIVPHGDHYIIPKNEISASELAAAEAFLSGRGNSRTRYRONSQNTSRTKWP 239
Db      179 TGDAYIVPHGDHYIIPKNEISASELAAAEAYWNG-----KQSRPSSSSSYNA 227
QY      240 SVSNPGTNTNTSNNSNTSGASQNDISLKLQYKPLISQHVESGLVFPDPAQITSR 299
Db      228 NPAQPRUSENHLVTTPYHNO-NGENISSLRRLVAKPLSERHVESGLVFPDPAQITSR 286
QY      300 TARGAAPHGHDHYHFIPIYQMSLEERLARIIPRYRNSHWVPDRPEQSPQPTPEPSP 359
Db      287 TARGAAPHGHDHYHFIPIYQMSLEERLARIIPRYRNSHWVPDRPEQSPQPTPEPSP 346
QY      360 GPQAPNLK-IDSN---SLYSQLVKRYGSEYVEEKGISRYVAKDIPSTYKNLESKL 415
Db      347 SPOAPNPQAPASNPIDKLVKEAVRKVGDDYVEEENGVSRYIPAKDLSARTAGISKL 406
QY      416 SKQSVSHTLTAKKENVAPROEFYDRAVNLTAHAKLFPNKRNSPFDLKLLESLN 475
Db      407 AKQBSLSHKLGAKTDLPSDRREFYKAYDLAIHODLDNKRQYDFEALDLLESLK 466
QY      476 DESTNKEKLVDDLAFAPITHPERLGPNSQIEYTEDEVARIAQLADYTTSDGYIPDEH 535
Db      467 DVXSDKVLVXDLAFAPIRHPRILGKPNQIITYTDEIVAKLACKYTTEDGYIPDR 526
QY      536 DIISDEGDAYTTPMGSHWIGKDSLDKQKVAQAAYTKKGIILPSPDADYKANPTGDS 595.
Db      527 DITSDEGDAYTTPMTSHWIKDLSLSEAEPAQAAYAKEGILTPPSTPDHODSGNTEAKG 586
QY      596 AAIYINRYKGEKRIPLVRLPYVHEHTEVKNGLIIPKHDYHNIKFAPFDHYKAPNG 655
Db      587 ABAIYNRYKAKKAVPLDRMPYNIQYVEVKNGSLIIPHYDHNHNIKEFWEDEGLYEAPK 646
QY      656 YTLDELFIATIKYVYEHDPDRPHSNDGNGNASEHVLGKKHSEDPNKNFKADEE-----P 709
Db      647 YTLDELFIATIKYVYEHDPDRPHSNDGNGNADHVRKKNKADODSKPDEDKGHDVESEPTHE 706
QY      710 VEET-----PAPEVPOVETEKVEAOLKEAEVLAKYTD 743
Db      707 EETPREKQSEKPEKPEKPEEPESEPEPEVETEKVEKLRLEADLLGRIOD 763

RESULT 35
ABP54589
ID      ABP54589 standard; Protein; 763 AA.
XX      ABP54589;
AC      ABP54589;
XX      04-SEP-2002 (first entry)
DT      S. pneumoniae SP0042 protein sequence SEQ ID NO: 66.
DE      XX
```

KW	Streptococcus pneumoniae; epitope; vaccine; antigenic protein;	
KW	antibacterial; Streptococcal infection; detection.	
OS	Streptococcus pneumoniae.	
XX	US2002061545-A1.	
XX	23-MAY-2002.	
XX	22-JAN-2001; 2001US-0765272.	
XX	30-OCT-1997; 97US-0961083.	
XX	(CHOI/) CHOI G H.	
XX	(KUNS/) KUNSCH C A.	
XX	(BARA/) BARASH S C.	
XX	(DILL/) DILLON P J.	
XX	(DOUG/) DOUGHERTY B.	
XX	(FANN/) FANNON M R.	
XX	(ROSE/) ROSEN C A.	
PI	Choi GH, Kunsch CA, Barash SC, Dillon PJ, Dougherty B, Fannon MR;	
FI	Rosen CA;	
XX	WPI; 2002-479261/51.	
DR	N-PSDB; ABQ84824.	
XX	New Streptococcus pneumoniae antigens, useful for detecting	
PT	Streptococcus and for preventing or attenuating disease caused by	
PT	Streptococcus infection -	
XX	Claim 11; Page 29; 70pp; English.	
PS	ABQ84792 to ABQ84904 represents nucleic acids which encode the	
XX	Streptococcus pneumoniae antigens given in ABP54557 to ABP54669.	
CC	The S. pneumoniae antigens have antibacterial activity and can be	
CC	used in vaccines. The S. pneumoniae antigens can also be used to	
CC	prevent or attenuate a Streptococcal infection in an animal. The	
CC	polynucleotides encoding the S. pneumoniae antigens can be used to	
CC	detect Streptococcus nucleic acids. ABQ84905 to ABQ85130 represent	
CC	primers used in the cloning of S. pneumoniae ORFs (open reading frames)	
CC	which are used in an example from the present invention.	
CC	Sequence 763 AA;	
SQ		
Query Match	63.6%; Score 2649.5; DB 23; Length 763;	
Best Local Similarity	66.8%; Pred. No. 3.9e-187;	
Matches 519; Conservative	77; Mismatches 132; Indels 49; Gaps 8;	
OY	1 SYELGLYAFRTV-KENNRRVSYIDGKQATOKFENLTPDEVSKEGGINABOIVYKIDGGYV	59
DB	2 SYELGRHQAQGVKKESNRVSYIDGQAQKAKENLTPDEVSKREGINAAQXVYKIDGGYV	61
OY	60 TSHGDHYHYNGKVPYDAIISEELMKDPNYKLKDEDIVNEVKGGVYIKVDGKYVYLKD	119
DB	62 TSHGDHYHYNGKVPYDAIISEELMKDPNYQLKSDIVNEIKGGYVIAKNGKYVYLKD	121
OY	120 AAHADNVPTKEINQKQEHQSHRGCGTTPRDGAALALARSQGRYTTDGYTFNADIIED	179
DB	122 AAHADNVPTKEIKRKQKQERSHNHNS--RADNVAALARAQGRYTTDGYTFNADIIED	178
OY	180 TGDAYIVVPGHGHYHIIPKNELSASELAALAAFLSGRGLINSRTYRQRNSDSTRTNWVP	239
DB	179 TGDAYIVVPGHGHYHIIPKNELSASGLAALAAEAVWNG-----KQGRSSSSSSYNA	227
OY	240 SVSNPCTTNVITSNNSNTNSQAQSGNDIDSLIKQLYKPLPSGRHVESGLVDFPAQITSR	299
DB	228 NPAQRISENHNLTVTPYHQ-NQGENISISLIRELYARPLSRHHVESGLIFDPAQITSR	286
OY	300 TARGVAVHGHHYHIIPTYSOMSELEERTIARIILPLYRNNHWVPSDRPQSPQPTPEESP	359
DB	287 TARGVAVHGHHYHIIPTYSOMSELEKRIARIILPLYRNSHWVPSDRPQSPQPTPEESP	346

OY	360	GPQAPNKK-IDSN---	SLYSQALRXKQCEGVFEFEKGISRVAFMKDLSSEYTKNLESKL	415
OY	360	GPQAPNKK-IDSN---	SLYSQALRXKQCEGVFEFEKGISRVAFMKDLSSEYTKNLESKL	415
DB	347	SPQAPNPQAPASNIDEKLVEAVRKVGDDGVFEENGVSRYIPAKDLISAETAAIGIDSKL	406	
OY	416	SKQESVSHTLTAKKENVAAPRODFYDKAVNLLTEAHKALFXNKGNSDFQALDKLERLN	475	
DB	407	AKQBSLSHKLGAQKTDLPSSDREFYFNKAYDLARIHODLNNKQOVDFEALDNLRLK	466	
OY	476	DESTNKEKLVDDLAFIAPITHPERLGRKNSQIEYTEDEVRIAGLADKYTTSDGYTFDEH	535	
DB	467	DVXSPKVLKXVDILAFIAPIRHPEELGRKNQAITYTDDIEIQAKLAGKTYTDEGYTFDPR	526	
OY	536	DIISPEGAYTTPHNGHSHWIGKDSLSDKEKTAQAAYTTEKSLTPSPSPADVKAAPTGS	555	
DB	527	DITSDEGAYYTPHNTSHMWIKKDSLSEAPERAQAAYAKKELTPPSTHDQSGNTEAKG	586	
OY	596	AAAIYNRKGEKRIPLVRLPYWVEVTEVKGNGLIIPKHDYHNTKFAWFDHTYKAPNG	655	
DB	587	AEAIYNRKAAKAVLDMFPYKQITVEYKNGSLIIPHYDHNHKKFEMFDEGLYEAAPG	646	
OY	656	YTLIEDLFAITIKYYVEHPDERPHSNDGWNASEHVLGKKDHSDEPNKNFKADEP-----P	709	
DB	647	YTLIEDLTAIVKYVHEPHNERPHSDNGFNGASDHVGRNKGQADNTQTEKPESEKPOTEKR	706	
OY	710	VEET-----	PAEPVPQVETIEKVNAQLKEAVLLAKVTD	743
DB	707	EEETPREKPOSEKPESEPKPTPEPESEPESEEPQVETEKVEKLEAREADLLGKIOD	763	
RESULT 36				
ID	AA05753	standard; Protein; 721 AA.		
AC	AA05753;			
XX	19-JUL-1999	(first entry)		
DE	Streptococcus pneumoniae 79	kda complement C3-degrading protease.		
KW	Human complement C3-degrading protease; Vaccines; Infection; meningitis; pneumonia; xerograft transplantation; transplant rejection; inflammation.			
OS	Streptococcus pneumoniae.			
XX	Key	Location/Qualifiers		
FT	Protein	170..227		
FT	Protein	/note="claimed protein of Claim 14"		
FT	Protein	258..300		
XX	PM	/note="claimed protein of Claim 15"		
XX	PM	WO915675-A1.		
XX	01-APR-1999.			
XX	24-SEP-1998;	98WO-US20186.		
XX	24-SEP-1997;	97US-005907.		
XX	(AMCY) AMERICAN CYANAMID CO.			
XX	(MINU) UNIV MINNESOTA.			
XX	Cheng Q, Finkel DJ, Green BA, Hostetler MK, Masi AM;			
XX	WPI, 1999-254719/21.			
XX	N-PSDB; AAX25394.			
XX	New isolated human complement C3-degrading proteinase			
XX	Claim 59; Page 54-58; 66pp; English.			
CC	The present sequence is a 79 kda protein of Streptococcus pneumoniae serotype 4 that is capable of degrading human complement protein C3			

(HCPC3). The sequence was deduced from the open reading frame of an isolated gene (see AX25394). A smaller (20 kDa) HCPC3 protease is encoded by an open reading frame of *S. pneumoniae* serotype 3 and this open reading frame may be part of a larger open reading frame corresponding to the present sequence from the *S. pneumoniae* serotype 4 genome. Amino acids 1-58 and 90-132 of the 20 kDa protein have substantial sequence identity with amino acids 170-227 and 258-300 of the 79 kDa protein. Proteins and peptides or polypeptides containing these regions, and DNA sequences encoding them, are claimed. HCPC3 proteases and polypeptides can be used as immune system stimulating compositions (claimed). They can produce an immune response against *S. pneumoniae* to immunize or treat a mammalian subject against infection or colonization (claimed). They can produce a B cell response, a T cell response, an epithelial cell response, or an endothelial cell response (claimed). The expression of the proteins on the surface of an organ of an animal used in xenotransplantation can be used to inhibit C3-mediated inflammation and rejection.

Sequence 721 AA:

Query Match 58.4%; Score 2431.5; DB 20; Length 721;
 Best Local Similarity 64.9%; Pred. No. 56-171;
 Matches 476; Conservative 79; Mismatches 132; Indels 47; Gaps 7;

85 MKDPYKLDKEDIVNEVKGYYIKVGGKYYVVKDAADNRTEKEINROKQESHORE 144
 1 MKDPYKLDKEDIVNEVKGYYIKVGGKYYVVKDAADNRTEKEINROKQESHORE 60
 145 GGTPRNDGVALARSQGRYTTDDGYIFNADIIEDTDGAYIVPHGDHYHIFPKNELSASE 204
 61 GGS--NDQAVVVARAQGRYTTDDGYIFNADIIEDTDGAYIVPHGDHYHIFPKNELSASE 118
 205 LAAAFAPLSGRNLSNRTRYRQNDNTRTWNVSVPSTGTTNTNTSNTNSQASQS 264
 119 LAAAFAYWNG-----KQGRPSSSSYNNPAPRLSEHNNLTVPPTYHQ--NQG 166
 265 NDIDLKQLYKLPISORHVESDGLVFPDPAQITSTARGAVAPHGDHYHIFYSQMSLE 324
 167 ENISLLELAKPLSERHVESDGLVFPDPAQITSTARGAVAPHGDHYHIFYSQMSLE 226
 325 ERIARIPLRYRSHNWVPDSRPEQSPQPTPEPSGQAPAPNLK-IDSN--SSLVSQLV 380
 227 KRIARIPLRYRSHNWVPDSRPEQSPQPTPEPSGQAPAPNPQAPNPIDKLVKEAV 286
 331 RRVGGYVFEKIGSRVYFAKDLPSFTVKLSKLSKQSVSHITLTAKENAVPRDOEFY 440
 287 RRVGGYVFEENGVSRYIPAKDLAETAGIDSKLAKOESLSHKLGAKTDLPSSDREY 346
 441 DKAYVLTLEAHKALFXNKGNSDFQALDKLERLNDSENGKELVDDLLAFLAPITPHER 500
 347 NKAYDLARIHODLLDNKGRQVDFEALDNLRLKDVPSDKLVDDLLAFLAPITPHER 406
 501 LGKPNQLEYTEDEVRVIAQLADKYTTSDGYIFDEHDIISDEGDVYVTPHMGSHWIGDS 560
 407 LGKPNQLEYTEDEVRVIAQLADKYTTSDGYIFDEHDIISDEGDVYVTPHMGSHWIGDS 466
 561 LSDKKVAAQATTKKGLIPSPDADVKANPTGDSAAAIYRNVKKEKRIPLVRLYMEH 620
 467 LSEARAAQAAYAKKKGITPSTHDQSGNTEAKGAEALYRNVKKAKKVPLDRMYNIQY 526
 621 TVEVNGKGLIIPKHGHYHNIKFAFPDHTYKAPNGTLEDFATIKYYVHDPDEPHNSD 680
 527 TVEVNGKGLIIPKHGHYHNIKFAFPDHTYKAPNGTLEDFATIKYYVHDPDEPHNSD 586
 681 GWGNASEHVLGKKDSEDPNKNFKADEE-----FVEET----- 713
 587 GFGNNSDVHQRKNQADQNTQTEKPESEKQTEKEEETPREKQSEKPEPKTEPE 646
 714 --PAPEVQVETEVKAEVLLAKVTDDSLKANATEFLAGLRNNLTQINDNSI 771
 647 ESPSESEPOVETKEVEKTLREAEADLQKIQPIIKSNKAKETLTGLKNLLFGTDNNYI 706

Qy 772 MAEAKLALIKGS 785
 Db 707 MAEAKLALIKGS 720

RESULT 37
 AAB12730
 ID AAB12730 standard; Protein; 613 AA.

AC AAB12730;

DT 21-NOV-2000 (first entry)

DE Streptococcus pneumoniae BVH-11C protein antigen SEQ ID NO:63.

KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;

OS otitis media; pneumonia; immunisation; bactericidal.

OS Streptococcus pneumoniae.

PN WO200039299-A2.

PD 06-JUL-2000.

PF 20-DEC-1999; 99WO-CA01218.

PR 23-DEC-1998; 98US-0113800.

PA (BIOC-) BIOCHEM PHARMA INC.

PI Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

DR WPI; 2000-452397/39.

PT Streptococcal antigens useful for vaccinating against e.g. meningitis,

PS otitis media, bacteraemia and/or pneumonia -

PS Claim 18; Fig 28; 106pp: English.

CC The present invention describes nucleic acids (I) encoding protein

CC antigens (II) from Streptococcus pneumoniae. The protein antigens

CC have bactericidal activity. The nucleic acids, encoding the protein

CC antigens, may be used for the recombinant production of the proteins

CC they encode. The protein antigens may then be used as vaccines for the

CC prevention and treatment of Streptococcal infections in mammals

CC (especially humans) which result in, e.g. meningitis, otitis media,

CC bacteraemia and/or pneumonia. The present sequence represents the

CC *S. pneumoniae* BVH-11C protein antigen.

CC Sequence 613 AA;

Qy 208 AEAFLSGRGNLSNRTRYRQNDNTRTWNVSVPSTGTTNTNTSNTNSQASQNDI 267

Db 1 AEAFLSGRGNLSNRTRYRQNDNTRTWNVSVPSTGTTNTNTSNTNSQASQNDI 60

Qy 268 DSLKQLYKLPISORHVESDGLVFPDPAQITSTARGAVAPHGDHYHIFYSQMSLEBRI 327

Db 61 DSLKQLYKLPISORHVESDGLVFPDPAQITSTARGAVAPHGDHYHIFYSQMSLEBRI 120

Qy 328 ARRIPLRYRSHNWVPDSRPEQSPQPTPEPSGQAPAPNLK-IDSN--SSLVSQLVKV 383

Db 121 ARRIPLRYRSHNWVPDSRPEQSPQPTPEPSGQAPAPNPQAPNPIDKLVKEAVKV 180

Qy 384 GEGYVFEKIGSRVYFAKDLPSFTVKLSKLSKQSVSHITLTAKENAVPRDOEFYDK 443

Db 181 GEGYVFEENGVSRYIPAKDLAETAGIDSKLAKOESLSHKLGAKTDLPSSDREYVKA 240

Qy 444 YNLTLEAHKALFXNKGNSDFQALDKLERLNDSENGKELVDDLLAFLAPITPHERLKG 503

```

Db      241 YDLARIHODLDNKRQVDFEALDNLRLKQVSSDKVTLVDDILAFLAPIRHPRLGK 300
QY      504 PMSQIETDEVEVRIADLADKYTSDGYIFDEHDIISDEGAYVTPHMGSHWIGKDSLSD 563
Db      301 PNAQIITYTDDDEIOVAKLAGKTYTDEGYIFDPRIITSDGAYVTPHMTSHWIKKDSLSE 360
QY      564 KEKVAAQAYTKEKGIPLPSPDADVKANPTGDSAAAIYNRVKGSKRIPLVRLPYMVEHTYE 623
Db      361 AERAAQAAYAKKEKGLPPSTHDQDSGTEAKGAELIYNRKAAKVPDLDMRYNLQYVYE 420
QY      624 VKNGNLIIPKDHYNHNIKAFMDPDHYKAPNGTTLDELFATIKYYVHEHDERPHSDGWC 663
Db      421 VKNGSLIIPHYDHYHNHNIKFEWPEDEGLYEAPKGYTLLEDLATVKKYVEHPNERPHSDNGFG 480
QY      684 NASEHVLGKKHSDSDPNKNFKAD-----EEPEEVT 713
Db      481 NASDHVQRNKGADJNQTKEPSEKQPTKEPEEETPREKPOSEKPESEKPTPEEPEES 540
QY      714 PAPEVPOVETKEVEAQLKEAEVLLAKVTDSLSKANATETLAGLRNNLTLOIMDNNSIMA 773
Db      541 PESEEPQVETKEVEEKLREAEIDLGIQDPIIKSNAKETLTGLKNLTLFGTDNNTIMA 600
QY      774 EAEKLLALLKGS 785
Db      601 EAEKLLALLKES 612

RESULT 38
AAU84029
ID      AAU84029 standard; Peptide; 613 AA.
AC      AAU84029;
XX      XX
XX      08-MAY-2002 (first entry)
DE      Truncated variant of S. pneumoniae BVH-11, BVH- version #2.
XX      XX
XX      BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KM      pneumonia; streptococcal bacterial infection; mutant; mutain.
XX      XX
OS      Streptococcus pneumoniae.
OS      Synthetic.
XX      XX
PN      MO200198334-A2.
XX      XX
PD      27-DEC-2001.
XX      XX
PF      19-JUN-2001; 2001WO-CA00908.
XX      XX
PR      20-JUN-2000; 2000US-212683P.
XX      XX
PA      (SHIR-) SHIRE BIOCHEM INC.
XX      XX
PI      Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX      MPI; 2002-122272/16.
XX      XX
PT      New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT      epitope-bearing polypeptides, useful as vaccine components for treating
PT      or preventing streptococcal infections such as otitis media,
PT      meningitis, and bacteraemia
XX      XX
PS      Example 1; Page -: 113pp; English.
XX      XX
CC      The invention describes an isolated polypeptide (I) with 70-90%
CC      identity to Streptococcus pneumoniae protein BVH-3, BVH-11, variants of
CC      BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC      comprising (I) is useful for therapeutic or prophylactic treatment of
CC      meningitis, otitis media, bacteraemia or pneumonia infection in an
CC      individual susceptible to these disorders. (II) is also useful for
CC      therapeutic or prophylactic treatment of any streptococcal bacterial
CC      infection (e.g., caused by Streptococcus pneumoniae, group A

```

```

CC      Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC      as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novacida or
CC      Streptococcus aureus) in an individual susceptible to the infection.
CC      A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC      techniques. The Streptococcus polypeptides are useful in a diagnostic
CC      test for S. pneumoniae infection. (III) is useful for designing DNA
CC      probes for use in detecting the presence of Streptococcus in a biological
CC      sample suspected of containing the bacteria. The DNA probes may also be
CC      used for detecting circulating S. pneumonia nucleic acid in a sample for
CC      diagnosing streptococcal infections. This sequence represents a truncate
CC      of a Streptococcus pneumoniae gene used to obtain antigenic peptides.
CC      Described in the method of the invention.
CC      Note: This sequence does not appear in the specification but has
CC      been created according to information given in the invention.
XX      XX
SQ      Sequence      613 AA;
Query Match      52.1%; Score 2170; DB 23; Length 613;
Best Local Similarity 68.5%; Pred. No. 8,9e-152;
Matches 419; Conservative 59; Mismatches 100; Indels 34; Gaps 3;
QY      208 AEAFLSGRNLNSRFRYRONSNDNTRTWVSNSPCTTNTNTSNTNSQAOSQNDI 267
Db      1 AEAFLSGRENLSNLTFRQNSDNTPRTNWVSNSPCTTNTNTSNTNSQAOSQNDI 60
QY      268 DSLKQLYKLPSQRHVESDGLVFPDPAQITSTARGAAPHGDHYFIYSQWSELEERI 327
Db      61 DSLKQLYKLPLSQRHVESDGLIFDPAQITSTARGAAPHGNHFIIFYEOMSELEEKI 120
QY      328 ARIIPLRYSNMHWPDSPREOPSPOPTPEPSGPOPAPLK-IDSN---SSLVSQLVRY 383
Db      121 ARIIPLRYSNMHWPDSPREOPSPOPTPEPSGPOPAAPNPQAPSPIDIKLVEKAVRY 180
QY      384 GEGYVEEEKIGIRYPAKDLPSFTYKNLESKSKQSVSHITLAKKENAAPPDQEFYDGA 443
Db      181 GDGYVEENGVSRYIPAKNLSAETAAGIDSKLAKOESLSHKLGAKKTDLPSSDREYENYA 240
QY      444 YNLTFAHAKLFPYKNKGRNSDFOALDKLRLNDESTNKEKLVDDILAFLAPIRHPRLG 503
Db      241 YDLARIHODLDNKRQVDFEALDNLRLKQVSSDKVTLVDDILAFLAPIRHPRLGK 300
QY      504 PMSQIETDEVEVRIADLADKYTSDGYIFDEHDIISDEGAYVTPHMGSHWIGKDSLSD 563
Db      301 PNAQIITYTDDDEIOVAKLAGKTYTDEGYIFDPRIITSDGDAYVTPHMTSHWIKKDSLSE 360
QY      564 KEKVAAQAYTKEKGIPLPSPDADVKANPTGDSAAAIYNRVKGSKRIPLVRLPYMVEHTYE 623
Db      361 AERAAQAAYAKKEKGLPPSTHDQDSGTEAKGAELIYNRKAAKVPDLDMRYNLQYVYE 420
QY      624 VKNGNLIIPKDHYNHNIKAFMDPDHYKAPNGTTLDELFATIKYYVHEHDERPHSDGWC 683
Db      421 VKNGSLIIPHYDHYHNHNIKFEWPEDEGLYEAPKGYTLLEDLATVKKYVEHPNERPHSDNGFG 480
QY      684 NASEHVLGKKHSDSDPNKNFKAD-----EEPEEVT 713
Db      481 NASDHVQRNKGADJNQTKEPSEKQPTKEPEEETPREKPOSEKPESEKPTPEEPEES 540
QY      714 PAPEVPOVETKEVEAQLKEAEVLLAKVTDSLSKANATETLAGLRNNLTLOIMDNNSIMA 773
Db      541 PESEEPQVETKEVEEKLREAEIDLGIQDPIIKSNAKETLTGLKNLTLFGTDNNTIMA 600
QY      774 EAEKLLALLKGS 785
Db      601 EAEKLLALLKES 612

RESULT 39
AAU84058
ID      AAU84058 standard; Peptide; 1126 AA.
AC      AAU84058;
XX      XX
XX      08-MAY-2002 (first entry)

```

XX S. pneumoniae derived chimeric peptide, NEW31.
 DE BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 XX pneumonia; streptococcal bacterial infection; mutant; mutein;
 KM BVH-11-2.
 XX Streptococcus pneumoniae.
 OS Synthetic.
 XX WO200198334-A2.
 PN 27-DEC-2001.
 XX 19-JUN-2001; 2001WO-CA00908.
 PF 20-JUN-2000; 2000US-212683P.
 XX (SHIR-) SHIRE BIOCHEM INC.
 PA Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 PI WPI; 2002-122272/16.
 DR New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 XX epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia
 XX
 XX Example 1; Page -: 113pp; English.
 PS The invention describes an isolated polypeptide (I) with 70-90%
 XX identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardiae or
 CC Streptococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a chimeric
 CC gene created from fragments and variant fragments of Streptococcus
 CC pneumoniae genes, described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.
 CC
 SQ Sequence 1126 AA;
 Query Match 45.6%; Score 1898; DB 23; Length 1126;
 Best Local Similarity 64.9%; Pred. No. 3,2e-131;
 Matches 366; Conservative 63; Mismatches 101; Indels 34; Gaps 3;
 QY 266 DIDLKLYLPLSQRHVESDGLVDPDPAQITSRPARGAAYPHGHYFIPISQMSLEE 325
 DB 2 DIDLKLYLPLSQRHVESDGLVDPDPAQITSRPARGAAYPHGHYFIPISQMSLEE 61
 QY 326 RIARIIPRYSNHWPDSRPEPSPOPTPEPSPOQAPNLK-IDSN---SILVSQILVR 381
 DB 62 RIARIIPRYSNHWPDSRPEPSPOPTPEPSPOQAPNLK-IDSN---SILVSQILVR 121
 QY 382 KVGEGVVEEKIGSYVFAKLPSTVKNLESKLSKSSVSHTLTAKKENANPAPQDEYD 441
 DB 122 KVGEGVVEEKIGSYVFAKLPSTVKNLESKLSKSSVSHTLTAKKENANPAPQDEYD 181
 QY 442 KAYNLTBAHAKLFXNKGNSDFQALDKLERLNDESINKEKLVVDLLAFLAPITHPRL 501

DB 182 KAYNLTBAHAKLFXNKGNSDFQALDKLERLNDESINKEKLVVDLLAFLAPITHPRL 241
 QY 502 GKPSQIEYTEDDEVRIAQLADKTYTSDGYIFDEHDIISDEBDAVTPRMGSHNIGDSL 561
 DB 242 GKPSQIEYTEDDEVRIAQLADKTYTSDGYIFDEHDIISDEBDAVTPRMGSHNIGDSL 301
 QY 562 SDKEKVAQAAYTKKEGILPSPSPADVAKNPTGDSAAAAYNNVKEGKRIPLVRLPYMVEHT 621
 DB 302 SEABRAAQAAYTKKEGILPSPSPADVAKNPTGDSAAAAYNNVKEGKRIPLVRLPYMVEHT 361
 QY 622 VEANGNLIIPKDHYNHNIKFPWPDHTTYXAPNGYTTLEDLPATIKYVYEHDESPHSNDG 681
 DB 362 VEANGNLIIPKDHYNHNIKFPWPDHTTYXAPNGYTTLEDLPATIKYVYEHDESPHSNDG 421
 QY 682 WGNASRYVLGKKHSDSPNKKFKAD-----EEPYE 711
 DB 422 FGNSADHVQRNKNQADTNQTEKPESEKPKTEKEEETPREKPKQSEKSPKTEPEE 481
 QY 712 ETPAPEVPOVETEKVEAQLKEAEVLAKVTDSLSKANATETLAGLRNNLTLOQMDNSI 771
 DB 482 ETPAPEVPOVETEKVEAQLKEAEVLAKVTDSLSKANATETLAGLRNNLTLOQMDNSI 541
 QY 772 MABAELKLLALKGSPSSVSKETI 795
 DB 542 MABAELKLLALKGSPSSVSKETI 565
 RESULT 40
 AAU84057
 ID AAU84057 standard; Peptide; 1365 AA.
 AC AAU84057;
 XX 08-MAY-2002 (first entry)
 DT S. pneumoniae derived chimeric peptide, NEW30.
 XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KM pneumonia; streptococcal bacterial infection; mutant; mutein;
 KM BVH-11-2.
 XX Streptococcus pneumoniae.
 OS Synthetic.
 XX WO200198334-A2.
 PN 27-DEC-2001.
 XX 19-JUN-2001; 2001WO-CA00908.
 PF 20-JUN-2000; 2000US-212683P.
 XX (SHIR-) SHIRE BIOCHEM INC.
 PA Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 PI WPI; 2002-122272/16.
 DR New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 XX epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia
 XX
 XX Example 1; Page -: 113pp; English.
 PS The invention describes an isolated polypeptide (I) with 70-90%
 XX identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial

infection (e.g., caused by *Streptococcus pneumoniae*, group A *Streptococcus* such as *Streptococcus pyogenes*, group B *Streptococcus* such as *Streptococcus agalactiae*, *S. dysgalactiae*, *S. uberis*, *S. novae* or *Staphylococcus aureus*) in an individual susceptible to the infection. A polynucleotide (III) encoding (I) is useful in DNA immunisation techniques. The *Streptococcus* polypeptides are useful in a diagnostic test for *S. pneumoniae* infection. (III) is useful for designing DNA probes for use in detecting the presence of *Streptococcus* in a biological sample suspected of containing the bacteria. The DNA probes may also be used for detecting circulating *S. pneumoniae* nucleic acid in a sample for diagnosing streptococcal infections. This sequence represents a chimeric gene created from fragments and variant fragments of *Streptococcus pneumoniae* genes, described in the method of the invention. Note: This sequence does not appear in the specification but has been created according to information given in the invention.

Sequence 1365 AA;

Query Match 45.4%; Score 1891.5; DB 23; Length 1365;
Best Local Similarity 65.0%; Pred. No. 1.3e-130;
Matches 367; Conservative 60; Mismatches 99; Indels 39; Gaps 4;

QY 266 DIDSLLKOLYKPLSGRHVESDGLVFPDPAQITSRTARGAVAPHGHHYPIYSOMSELEB 325
DB 2 DIDSLLKOLYKPLSGRHVESDGLVFPDPAQITSRTARGAVAPHGHHYPIYSOMSELEB 61
QY RIARIIPLRYSNHWVPSRPPSPQPTPEPSPQPPAPNPK-IDSN---SLVSQVLR 381
DB RIARIIPLRYSNHWVPSRPPSPQPTPEPSPQPPAPNPK-IDSN---SLVSQVLR 121
QY 382 KYGEGVFEPEKGISRYVFAKDLPSFTVKNLESLSKOSVSHTLTAKKENVAPRODFYD 441
DB 122 KYGEGVFEPEKGISRYVFAKDLPSFTVKNLESLSKOSVSHTLTAKKENVAPRODFYD 181
QY 442 KAYNLTFAHAKLFPNKGNSDFQALDKLERLNDESTNKEKLVDDLAFLAPITHPERL 501
DB 182 KAYNLTFAHAKLFPNKGNSDFQALDKLERLNDESTNKEKLVDDLAFLAPITHPERL 241
QY 502 GKPNQOIEYTEDVEVRIADLADKYTTSDGYIFDEHDIISDEGAYVYPHNGSHWIKGDSL 561
DB 242 GKPNQOIEYTEDVEVRIADLADKYTTSDGYIFDEHDIISDEGAYVYPHNGSHWIKGDSL 301
QY 562 SDKEVAAQAYTKKGLIPSPDADYKANPTGDSAAAIYNRVYGEKRIPLVLPYWEHT 621
DB 302 SEAEPAQAQAYAKKGLTPSTDHODSGTEAKGAEAIYNRVYKAQKVPDLDMFYNLOYT 361
QY 622 VEVKNGNLIIPKDHYNHIFKAMPDHTYKAPNGYTLLEDLFAITKYVEHPDERPHSNDG 681
DB 362 VEVKNGNLIIPKDHYNHIFKAMPDHTYKAPNGYTLLEDLFAITKYVEHPDERPHSNDG 421
QY 682 WGNASEHVLGKKDHSDDPNKFKAD-----EEPYE 711
DB 422 FGNASDHVQRNKGADTNQTEKPESEKQTEKPEEETPREEKPOSEKPEEPTPEEPE 481
QY 712 ETPAPPEVQVETKEAEVLAQAEVLAQAEVLAQAEVLAQAEVLAQAEVLAQAEVLAQAEV 771
DB 482 ETPAPPEVQVETKEAEVLAQAEVLAQAEVLAQAEVLAQAEVLAQAEVLAQAEVLAQAEV 541
QY 772 MAEAEKLLALLKGS-----NPSSVS 791
DB 542 MAEAEKLLALLKGS-----NPSSVS 566
RESULT 41
AAB12734 standard; Protein; 555 AA.
XX AAB12734;
AC AAB12734;
XX 21-NOV-2000 (first entry)
DE Streptococcus pneumoniae NEW4 protein antigen SEQ ID NO:67.
XX

KW Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;
KW prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;
KW otitis media; pneumonia; immunisation; bactericidal.
OS Streptococcus pneumoniae.

PN WO200039299-A2.

PD 06-JUL-2000.

PF 20-DEC-1999; 99WO-CA01218.

PR 23-DEC-1998; 98US-0113800.

XX (BIOC-) BIOCHEM PHARMA INC.

PA Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;

PI WPI; 2000-452397/39.

XX Streptococcal antigens useful for vaccinating against e.g. meningitis,
PT otitis media, bacteraemia and/or pneumonia -
XX

PS Claim 18; Fig 32; 106pp; English.

CC The present invention describes nucleic acids (I) encoding protein
CC antigens (II) from *Streptococcus pneumoniae*. The protein antigens
CC have bactericidal activity. The nucleic acids, encoding the protein
CC antigens, may be used for the recombinant production of the proteins
CC they encode. The protein antigens may then be used as vaccines for the
CC prevention and treatment of streptococcal infections in mammals
CC (especially humans) which result in, e.g. meningitis, otitis media,
CC bacteraemia and/or pneumonia. The present sequence represents the
CC *S. pneumoniae* NEW4 protein antigen.

XX Sequence 555 AA;

Query Match 45.3%; Score 1888; DB 21; Length 555;
Best Local Similarity 65.7%; Pred. No. 5.8e-131;
Matches 364; Conservative 59; Mismatches 97; Indels 34; Gaps 3;

QY 266 DIDSLLKOLYKPLSGRHVESDGLVFPDPAQITSRTARGAVAPHGHHYPIYSOMSELEB 325
DB 1 DIDSLLKOLYKPLSGRHVESDGLVFPDPAQITSRTARGAVAPHGHHYPIYSOMSELEB 60
QY 326 RIARIIPLRYSNHWVPSRPPSPQPTPEPSPQPPAPNPK-IDSN---SLVSQVLR 381
DB 61 RIARIIPLRYSNHWVPSRPPSPQPTPEPSPQPPAPNPK-IDSN---SLVSQVLR 120
QY 382 KYGEGVFEPEKGISRYVFAKDLPSFTVKNLESLSKOSVSHTLTAKKENVAPRODFYD 441
DB 121 KYGEGVFEPEKGISRYVFAKDLPSFTVKNLESLSKOSVSHTLTAKKENVAPRODFYD 180
QY 442 KAYNLTFAHAKLFPNKGNSDFQALDKLERLNDESTNKEKLVDDLAFLAPITHPERL 501
DB 181 KAYNLTFAHAKLFPNKGNSDFQALDKLERLNDESTNKEKLVDDLAFLAPITHPERL 240
QY 502 GKPNQOIEYTEDVEVRIADLADKYTTSDGYIFDEHDIISDEGAYVYPHNGSHWIKGDSL 561
DB 241 GKPNQOIEYTEDVEVRIADLADKYTTSDGYIFDEHDIISDEGAYVYPHNGSHWIKGDSL 300
QY 562 SDKEVAAQAYTKKGLIPSPDADYKANPTGDSAAAIYNRVYGEKRIPLVLPYWEHT 621
DB 301 SEAEPAQAQAYAKKGLTPSTDHODSGTEAKGAEAIYNRVYKAQKVPDLDMFYNLOYT 360
QY 622 VEVKNGNLIIPKDHYNHIFKAMPDHTYKAPNGYTLLEDLFAITKYVEHPDERPHSNDG 681
DB 361 VEVKNGNLIIPKDHYNHIFKAMPDHTYKAPNGYTLLEDLFAITKYVEHPDERPHSNDG 420
QY 682 WGNASEHVLGKKDHSDDPNKFKAD-----EEPYE 711
DB 421 FGNASDHVQRNKGADTNQTEKPESEKQTEKPEEETPREEKPOSEKPEEPTPEEPE 480

Qy 712 ETPAPEVPEVETKEVKAQLKEAVLLAKVTDSLSKANATETLAGLRNNLTLOIMDNNSI 771
 Db 481 ESPSESEPPQVETKEVKEKLREAEEDLKGIDPIIKSNAKETLTGKNLTFGTQDNNTI 540
 Qy 772 MAEAKLTALLKGS 785
 Db 541 MAEAKLTALLKES 554

RESULT 42

AAU84047 ID AAU84047 standard; Peptide; 555 AA.
 AC AAU84047;
 DT 08-MAY-2002 (first entry)
 DE Truncated variant of S. pneumoniae BVH-11, NEW4.
 KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mutein.
 OS Streptococcus pneumoniae.
 OS Synthetic.
 PN MO200198334-A2.
 PD 27-DEC-2001.
 PF 19-JUN-2001; 2001WO-CA00908.
 PR 20-JUN-2000; 2000US-212683P.
 PA (SHIR-) SHIRE BIOCHEM INC.
 PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 DR WPI; 2002-122272/16.
 XX
 PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia
 XX
 PS Example 1; Page -: 113pp; English.
 CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or
 CC Streptococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a truncate
 CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
 CC described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.
 XX
 SQ Sequence 555 AA;

Query Match 45.3%; Score 1888; DB 23; Length 555;
 Best Local Similarity 65.7%; Pred. No. 5.8e-131;

Matches 364; Conservative 59; Mismatches 97; Indels 34; Gaps 3;

Qy 266 DIDSLKQVLYKPLPSORHVESDGLVPDPAQITSTAGVAVPHGDHVFPIYSQMSLEEE 325
 Db 1 DIDSLKQVLYKPLPSORHVESDGLVPDPAQITSTAGVAVPHGDHVFPIYSQMSLEEK 60
 Qy 326 RIARILPLRYRSMHNVDPDSRPEQSPQPTPEPSPGPAPAPNLK-IDSN---SLVSQLVJR 381
 Db 61 RIARILPLRYRSMHNVDPDSRPEPSPQPTPEPSPGPAPAPNPIDELKVEAVR 120
 Qy 382 KVEGVYFEEKGIRRYPAKDLPEETVKNLESKLSKQESVSHLTAKKENVAPDQEFYD 441
 Db 121 KVGDGYYFEEENGVSRYIPAKVLSAETAGIDSKLAKQESLSHKLGAKKTDLPPSSDRFEYN 180
 Qy 442 KAVNLTAEAHKALFXNKGNSDPQALDKLERLNDESTNKEKLVDDLLAFAPITHEERL 501
 Db 181 KAYDRLARIHODLDNKGROVDFEALNLLERLKDVSDDKVLVDLLAFAPIRHEERL 240
 Qy 502 GKPNQIETEDEVEVRIQDLADKYTTSDDGYIFDEHDIISDEGDAYVTPHMGSHWIGDSL 561
 Db 241 GKPNQIETITDDELQVAKLAKGTYTTEGGYIFDPREDISDEGDAYVTPHMTSHWIKDSL 300
 Qy 562 SDKEKVAQAAYTKEKGLPPSPDADVKANPTGDSAAIYNRVKCEKRIPLVRLPYWVBHT 621
 Db 301 SEAEERAAQAAYAKEKGLTPSPSTDHQDSGNTAEAKGAERAIYNRVKAKKVPDLRMYNLYQT 360
 Qy 622 VEYKNGNLIIPKHYNINIKFAMFDDHTYKAPNGYTEDLFPATIKYVVEHPDEPHSNDG 681
 Db 361 VEYKNGSLIIPHYHYNIKFEMFDEGLYEAPKGYTLEDLLATVYVHEPNERPSDNG 420
 Qy 682 WGNASEHVLGKDKHSEDPNNKFNKAD-----EEPEVE 711
 Db 421 FGNASDHVQRNKNQADNTQTEKSESEKPOTEKEPEETPREKQSEKSPESPKTEPEEE 480
 Qy 712 ETPAPEVPEVETKEVKAQLKEAVLLAKVTDSLSKANATETLAGLRNNLTLOIMDNNSI 771
 Db 481 ESPSESEPPQVETKEVKEKLREAEEDLKGIDPIIKSNAKETLTGKNLTFGTQDNNTI 540
 Qy 772 MAEAKLTALLKGS 785
 Db 541 MAEAKLTALLKES 554

RESULT 43

AAU84031 ID AAU84031 standard; Peptide; 612 AA.
 AC AAU84031;
 DT 08-MAY-2002 (first entry)
 DE Truncated variant of S. pneumoniae BVH-11, NEW24.
 KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mutein.
 OS Streptococcus pneumoniae.
 OS Synthetic.
 PN MO200198334-A2.
 PD 27-DEC-2001.
 PF 19-JUN-2001; 2001WO-CA00908.
 PR 20-JUN-2000; 2000US-212683P.
 PA (SHIR-) SHIRE BIOCHEM INC.
 PI Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 DR WPI; 2002-122272/16.

PT New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
PT epitope-bearing polypeptides, useful as vaccine components for treating
PT or preventing streptococcal infections such as otitis media,
PT meningitis, and bacteraemia

XX Example 1; Page -: 113pp; English.

XX The invention describes an isolated polypeptide (I) with 70-90%
CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC comprising (I) is useful for therapeutic or prophylactic treatment of
CC meningitis, otitis media, bacteraemia or pneumonia infection in an
CC individual susceptible to these disorders. (II) is also useful for
CC therapeutic or prophylactic treatment of any streptococcal bacterial
CC infection (e.g., caused by Streptococcus pneumoniae, group A
CC Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nocardii or
CC Streptococcus aureus) in an individual susceptible to the infection.
CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC techniques. The Streptococcus polypeptides are useful in a diagnostic
CC test for S. pneumoniae infection. (III) is useful for designing DNA
CC probes for use in detecting the presence of Streptococcus in a biological
CC sample suspected of containing the bacteria. The DNA probes may also be
CC used for detecting circulating S. pneumonia nucleic acid in a sample for
CC diagnosing streptococcal infections. This sequence represents a truncate
CC of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
CC described in the method of the invention.
CC Note: This sequence does not appear in the specification but has
CC been created according to information given in the invention.

XX Sequence 612 AA;

Query Match 43.4%; Score 1808; DB 23; Length 612;

Best Local Similarity 58.0%; Pred. No. 5.7e-125;

Matches 357; Conservative 78; Mismatches 127; Indels 54; Gaps 7;

220 NSRTYRONSNTSRTNWV-PSVS-NGPTNTNTNSNSNTSQSQSNDISLLKOLYKL 277

5 NGKQGRSPSSSSSYNPNVQPRLENNHNLVPTTHQ-----NGENISLLRELKLVK 57

278 PLSQRVESDGLVFPDAQITSTRAGVAVPHGHYHPIFYPSQMSLEBRIARIIPIRYS 337

58 PLSERVESDGLVFPDAQITSTRAGVAVPHGHYHPIFYPSQMSLEBRIARIIPIRYS 117

338 NHVPSRPREPSPQPTPEPS-----PGQPAENLKIDNSNSLVQVLRKVEGVFEE 391

118 NHVPSRPREPSPQPTPEPS-----PGQPAENLKIDNSNSLVQVLRKVEGVFEE 175

392 KGISRYVFAKDLPESETVKNLESKLSQESVSHLTAKKENVAPRDOEFYDKAVNLTLEAH 451

176 NGVSRITPAKDLAEFTAGIDSKLAKQESLSHKLGAKTDLPSDSDEFYKAYDLARH 235

452 KALFANKGRNSDFQALDKLERLNDSSTNKEKLVDDLAFLADITPERIKGPNQIEYT 511

236 ODLLDKKGRQVDFEVDNLLERLKDVSSDKVLDLFLAFLIRPERIGKPNQIYTT 295

512 EDEVRITAOADKXTTDSGYIFDEHDIISDEGDAYVYVPHMCHSWIGDSIDEXYAAQA 571

296 DDEIQAALAGAKTTTDEGYIFDPRDITSDGDAYVYVPHMCHSWIKKDSISEKERAQA 355

572 YTEKESILPPSPADAVKAPFTGSAAYVYVGRGKRIPLVRLPYVWEHTVEYKGNLII 631

356 YAKEKLTPEPSTHOSGNTAEKAGABAIYKRVAAKVPFLDRMPYVLQYIVAYKNSLII 415

632 PKKHYNHIFKAFMDHTYKAPNGYLTLEDLFAITIKYVEHPDERPHSNDGMASEHVIG 691

416 PHYDHNHIFKAFMDHTYKAPNGYLTLEDLFAITIKYVEHPDERPHSNDGMASEHVIG 475

692 K-----KDHSE-----DPKNFPADEPEYBEP 714

476 NKADQDSKDEDEHDEVESEPTHPESDEKENHAGLNPASDNLVYKSTDETEEEDBT 535

715 ABPEVQVETEKVAQLKEAEVLLAKVYDLSLKANATETLAGLRNLLTIQIMDNNSIMAE 774

DB 536 DAEIPOVENSVINAKIAAEALKEVTPSPISRONAMETLTGKSLILGTROMNTISAE 595

QY 775 AEKLLALKGNSPSSV 790

DB 596 VDSLALLESQAPPI 611

RESULT 44

AAB12741

ID AAB12741 standard; Protein; 568 AA.

XX AAB12741;

AC AAB12741;

XX 21-NOV-2000 (first entry)

XX Streptococcus pneumoniae NEW10 protein antigen SEQ ID NO:74.

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

```

Db      180 YNKAYDLARIHODLLDNKRGVDFEVDNLRLKDVSSDKVLUDDLAFIARHPE 239
Qy      500 RLKGNISOLEYTEDEVRIALQADKYTTSDGYIFDEHDIISDEGDAYVPPHMGSHWIGKD 559
Db      240 RLKGNNAQIYTTDDEIQVAKLAGKTTTDEGYIFDPDITSDGDAYVTPHMTSHWIKD 299
Qy      560 SLSDKEKVAQAQYVTEKGIPLPSPPADYKANPTGDSAAAIYNRVKGEKRIPLVRLPYWE 619
Db      300 SLSEAEBAQAQYVAKKGLTPSTHQSNGTEAKGAELIYRVAAKVPIDRMFPYVLQ 359
Qy      620 HTVEKNGNLIIPKDHVHNIFKAMPDDHTYKAPNGYITLEDLFAITIKYVVEPDRPHSN 679
Db      360 YTVKNGSLIIPHYDHNHNIKFWPFBGLYEAPKGYSLBDLAVTKYVVEHPRERPHSD 419
Qy      680 DGMGNASEHVLGK-----KDHS-----DPKN 702
Db      420 NGFGNASHVRKNKADQDSKPEDKEHDEVESEPTHPESDEKENHAGLNPSADNLKXPSD 479
Qy      703 FKADEPEVEETPAPEVPQVETEKVQAQLEAEVLAKVTDSSLKANATETLAGLRNNLT 762
Db      480 TEETEEBAEDTTDEAEIPQVENSVINAKIADAEALLEKVTDPISRONMETLTGLKSSLL 539
Qy      763 LQIMDNNSIMAEKULALALKGSNPSV 790
Db      540 LGTKDNNTISAEDSLALALKEGAPAPI 567

RESULT 45
ID      AAU84088 standard; Peptide; 568 AA.
AC      AAU84088;
XX      08-MAY-2002 (first entry)
DE      Truncated variant of S. pneumoniae BVH-11-2, NEW10.
XX      BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW      pneumonia; streptococcal bacterial infection; mutant; mtein.
OS      Streptococcus pneumoniae.
XX      MO200198334-A2.
PN      27-DEC-2001.
XX      19-JUN-2001; 2001WO-CA00908.
PR      20-JUN-2000; 2000US-212683P.
XX      (SHIR-) SHIRE BIOCHEM INC.
PA      Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX      WPI; 2002-122272/16.
DR      New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX      epitope-bearing polypeptides, useful as vaccine components for treating
PT      or preventing streptococcal infections such as otitis media,
XX      meningitis, and bacteraemia
XX      Example 1; Page -; 113p; English.
CC      The invention describes an isolated polypeptide (I) with 70-90%
XX      identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
CC      BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
CC      comprising (I) is useful for therapeutic or prophylactic treatment of
CC      meningitis, otitis media, bacteraemia or pneumonia infection in an
CC      individual susceptible to these disorders. (II) is also useful for
CC      therapeutic or prophylactic treatment of any streptococcal bacterial
CC      infection (e.g., caused by Streptococcus pneumoniae, group A
CC      Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
CC      as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. nodocardia or

```

```

CC      Staphylococcus aureus) in an individual susceptible to the infection.
CC      A polynucleotide (III) encoding (I) is useful in DNA immunisation
CC      techniques. The Streptococcus polypeptides are useful in a diagnostic
CC      test for S. pneumoniae infection. (III) is useful for designing DNA
CC      probes for use in detecting the presence of Streptococcus in a biological
CC      sample suspected of containing the bacteria. The DNA probes may also be
CC      used for detecting circulating S. pneumonia nucleic acid in a sample for
CC      diagnosing streptococcal infections. This sequence represents a truncate
CC      of a Streptococcus pneumoniae gene used to obtain antigenic peptides,
CC      described in the method of the invention.
CC      Note: This sequence does not appear in the specification but has
CC      been created according to information given in the invention.
XX      SQ
XX      Sequence 568 AA:
XX
XX      Query Match 43.1%; Score 1795.5; DB 23; Length 568;
XX      Best Local Similarity 60.7%; Pred. No. 4,3e-124;
XX      Matches 345; Conservative 71; Mismatches 107; Indels 45; Gaps 4;
Qy      266 DIDLKOLYKLPQSORHVESDGLVFPDPAQITSTRAGVAVPHGDHYHFIYSQMSLEE 325
Db      2 NISSLRELYAKPLSERHVESDGLIFDPAQITSTRAGVAVPHGNHFIPIYEQMSLEK 61
Qy      326 RIARIIPLRYRSHWVPDSRPEQSPQPTPEPS-----PGQAPALUKIDSNSLSVQL 379
Db      62 RIARIIPLRYRSHWVPDSRPEQSPQPTPEPSLQAPAPNPAPNSPID--EKLVEA 119
Qy      380 VRKVGEGVFEKISIRYVFAKDLPSFTVKNLESKLSKQESVSHRTLAKKENVAPRQEF 439
Db      120 VRKVGDDGVFEENSVRIYPAKDLAETAGIDSKLAKQESLSHKLAKKTKDLPSSREF 179
Qy      440 YDKAYNLTAEAKLAPNKGKNSDFQALDKLERLNDESTKTEKLVLDLAFIAPITHE 499
Db      180 YNKAYDLARIHODLLDNKRGVDFEVDNLRLKDVSSDKVLUDDLAFIARHPE 239
Qy      500 RLKGNISOLEYTEDEVRIALQADKYTTSDGYIFDEHDIISDEGDAYVTPHMGSHWIGKD 559
Db      240 RLKGNNAQIYTTDDEIQVAKLAGKTTTDEGYIFDPDITSDGDAYVTPHMTSHWIKD 299
Qy      560 SLSDKEKVAQAQYVTEKGIPLPSPPADYKANPTGDSAAAIYNRVKGEKRIPLVRLPYWE 619
Db      300 SLSEAEBAQAQYVAKKGLTPSTHQSNGTEAKGAELIYRVAAKVPIDRMFPYVLQ 359
Qy      620 HTVEKNGNLIIPKDHVHNIFKAMPDDHTYKAPNGYITLEDLFAITIKYVVEPDRPHSN 679
Db      360 YTVKNGSLIIPHYDHNHNIKFWPFBGLYEAPKGYSLBDLAVTKYVVEHPRERPHSD 419
Qy      680 DGMGNASEHVLGK-----KDHS-----DPKN 702
Db      420 NGFGNASHVRKNKADQDSKPEDKEHDEVESEPTHPESDEKENHAGLNPSADNLKXPSD 479
Qy      703 FKADEPEVEETPAPEVPQVETEKVQAQLEAEVLAKVTDSSLKANATETLAGLRNNLT 762
Db      480 TEETEEBAEDTTDEAEIPQVENSVINAKIADAEALLEKVTDPISRONMETLTGLKSSLL 539
Qy      763 LQIMDNNSIMAEKULALALKGSNPSV 790
Db      540 LGTKDNNTISAEDSLALALKEGAPAPI 567

RESULT 46
ID      AAU84055 standard; Peptide; 1139 AA.
AC      AAU84055;
XX      08-MAY-2002 (first entry)
DE      S. pneumoniae derived chimeric peptide, NEW28.
XX      BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
KW      pneumonia; streptococcal bacterial infection; mutant; mtein;
KW      BVH-11-2.

```

XX Streptococcus pneumoniae.
 OS Synthetic.
 XX MO200198334-A2.
 XX 27-DEC-2001.
 XX 19-JUN-2001; 2001WO-CA00908.
 XX 20-JUN-2000; 2000US-212683P.
 XX (SHIR-) SHIRE BIOCHEM INC.
 XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX WPI; 2002-122272/16.
 DR
 XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia
 XX
 XX Example 1; Page -: 113pp; English.
 XX
 CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B, Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novardi or
 CC Streptococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic
 CC test for S. pneumoniae infection. (III) is useful for designing DNA
 CC probes for use in detecting the presence of Streptococcus in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating S. pneumonia nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a chimeric
 CC gene created from fragments and variant fragments of Streptococcus
 CC pneumoniae genes, described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.
 CC
 XX Sequence 1139 AA;
 SQ
 Query Match 43.1%; Score 1795.5; DB 23; Length 1139;
 Best Local Similarity 60.7%; Pred. No. 1.3e-123;
 Matches 345; Conservative 71; Mismatches 107; Indels 45; Gaps 4;
 QY 266 DIDLKQYKLPDLSQHVESDGLVDPDPAQTSRTARGAVPHGHHYPIPIPSOMSELEE 325
 DB 3 NISSLRREIYAKPLSRHVESDGLIFDPQITSRTRGAVPHGHHYPIPIPSOMSELEE 62
 QY 326 RIARIIPLRYRNHWVDSRPPSPDPTPEPS-----PGQPAENLKIDNSSLSVQL 379
 DB 63 RIARIIPLRYRNHWVDSRPPSPDPTPEPS-----PGQPAENLKIDNSSLSVQL 120
 QY 380 VAKVGEVYFEERKTSRYVFAKDLPSETYKNEKSKSRKESVSHLTAKENVARPDDEF 439
 DB 121 VAKVSDGVYFEENGVSRYIPADLSAETAGIDSKLAKQESLSHKLGAKKTLPSDRBF 180
 QY 440 YDKAVNLTEAKALFXKGRNSDPQALDKLERLNDESTNKEKLVDDLAFLAPITTHE 499
 DB 181 YKAKVDLARIHQDLDNKGROVDPEVDNLERLKDVSDDKXKLVDDLAFLAPITTHE 240
 QY 500 RLGKNSQIEYTEDVEVRIQAQADKYTSDGYIFDEHDIISDEGDVAVTBMHSHWIKGD 559
 DB 241 RLGKNAQITTYDDEIQVAKLAGKYTTEBDGYIFDPDITSDGDAVAVTBMHSHWIKGD 300

QY 560 SLSDKEKVAQAQYTKKGIPLPPSPADYKANPTGDSAAIYVRKGEKRIPLVRPLYWE 619
 DB 301 SLSEAFRAQAQYAEKELTPSTHDQSGNTEAGCAIYVRVAAKVPLDRPPLYNQ 360
 QY 620 HTVEVRKGNLLIIPKDHVNIKFAWFDHTYKAPNGYTLDELFAPIKYVHEHDERPHSN 679
 DB 361 YVEVRKGNLSLIIPHYDHNIKFEWFDEGLYEAAPKGYSLDELATVTKYVEHPNRPDSD 420
 QY 680 DGMGNASHVYLCK-----KDHE-----DPMKN 702
 DB 421 NGFGNASDHVRKNKADQSKPDEDKEHDEVESEPTHPESDEKENAGLNPASDNLKYPSTD 480
 QY 703 FKADEPVEETPAPEPVPQVETEKVEAQLKEAEVLLAKVTDSLKANETETLAGLRNLT 762
 DB 481 TEETEEAEDDTTDEEIRQVENSVINAKIADREALLEKVTDPSTRQNMETITGLKSSLL 540
 QY 763 LQIMDNSSIMAEKTLALLKGSNPSV 790
 DB 541 LGTKDNNTISAEDVSLALLKESQAPPI 568
 RESULT.47
 AAU84053 standard; peptide; 1378 AA.
 ID AAU84053
 XX
 AC AAU84053;
 XX
 DT 08-MAY-2002 (first entry)
 XX
 DE S. pneumoniae derived chimeric peptide, NEW26.
 XX
 KW BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
 KW pneumonia; streptococcal bacterial infection; mutant; mulein;
 KW BVH-11-2.
 XX
 OS Streptococcus pneumoniae.
 OS Synthetic.
 XX MO200198334-A2.
 XX 27-DEC-2001.
 XX 19-JUN-2001; 2001WO-CA00908.
 XX 20-JUN-2000; 2000US-212683P.
 XX (SHIR-) SHIRE BIOCHEM INC.
 XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
 XX WPI; 2002-122272/16.
 DR
 XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
 PT epitope-bearing polypeptides, useful as vaccine components for treating
 PT or preventing streptococcal infections such as otitis media,
 PT meningitis, and bacteraemia
 XX
 XX Example 1; Page -: 113pp; English.
 XX
 CC The invention describes an isolated polypeptide (I) with 70-90%
 CC identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
 CC BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
 CC comprising (I) is useful for therapeutic or prophylactic treatment of
 CC meningitis, otitis media, bacteraemia or pneumonia infection in an
 CC individual susceptible to these disorders. (II) is also useful for
 CC therapeutic or prophylactic treatment of any streptococcal bacterial
 CC infection (e.g., caused by Streptococcus pneumoniae, group A
 CC Streptococcus such as Streptococcus pyogenes, group B, Streptococcus such
 CC as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. novardi or
 CC Streptococcus aureus) in an individual susceptible to the infection.
 CC A polynucleotide (III) encoding (I) is useful in DNA immunisation
 CC techniques. The Streptococcus polypeptides are useful in a diagnostic

CC test for *S. pneumoniae* infection. (iii) is useful for designing DNA
 CC probes for use in detecting the presence of *Streptococcus* in a biological
 CC sample suspected of containing the bacteria. The DNA probes may also be
 CC used for detecting circulating *S. pneumoniae* nucleic acid in a sample for
 CC diagnosing streptococcal infections. This sequence represents a chimeric
 CC gene created from fragments and variant fragments of *Streptococcus*
 CC *pneumoniae* genes, described in the method of the invention.
 CC Note: This sequence does not appear in the specification but has
 CC been created according to information given in the invention.

XX Sequence 1378 AA;

Query Match 43.1%; Score 1795.5; DB 23; Length 1378;
 Best Local Similarity 60.7%; Pred. No. 1.7e-123;
 Matches 345; Conservative 71; Mismatches 107; Indels 45; Gaps 4;

QY 266 DIDLKQLYKPLPSQKRVESDGLVFPDAQTSRTARAGVAVHGHYHPIPSQMSLEE 325
 DB 3 NISSLRRLYAKPLSERHVESDGLIFDPAQTSRTARAGVAVHGHYHPIPSQMSLEE 62
 QY 326 RIARIIPRYSNHWPDSRPSPQSPPTPEPS-----PGQAPANKIDSNSSLSVQL 379
 DB 63 RIARIIPRYSNHWPDSRPSPQSPPTPEPSBQLPAPNPQAPNPID--EKLVAEA 120
 QY 380 VRKVEGVYFEKGISRYVPAKDLPSSETVKNLESKLSKQESVSHTLAKENVAPRDOEF 439
 DB 121 VRKVGDDGVFPEENGSRVYIPAKDLSAEIAGIDSKLAQESISHLGAKTDLPSDDEF 180
 QY 440 YDKAVNLTFAKALFKXKGRNSDFQALDKLERLNDSTNKEKLVDDLAFAPITPPE 499
 DB 181 YNKAVDLARIHQDLIDNKGQVDFEVLNLERLKDVS SDVKLVDDILAFAPIRHPE 240
 QY 500 RLGRNSQIEYTEDERIAQLADKYTTSQGYIFDEHDIISDEGDAYVPHGHSHWIKD 559
 DB 241 RLGRNAQITTTDDDEIQVAKLAGKTTEDGYIFEDRDIITSDGDAYVPHGHSHWIKD 300
 QY 560 SLSDKEKVAQAAYTTEKGIPLPSPDADYKANFTGDSAAIYVRVXGKRIPLVRLPYWE 619
 DB 301 SLSEERAAQAAYAKKGLTTPSTDHQDSNGTEAKGAELIYVRVAAKVPDLRNPVYLQ 360
 QY 620 HTVEKGNLIIIPKDHYNIKFAMFDHTYKAPNGYLTEDLPATIKYVHEPDERPSN 679
 DB 361 YTVKNGSLIIPHYDHYNIKFEWFDEGLYEAPKGSLEDLATVYVHEPDERPSD 420
 QY 680 DGMGNASEHVLGK-----KHSE-----DPKN 702
 DB 421 NGFGNASHVAKKAKADQSKDEDEKHEVSEPTHPESDEKENHAGLNPADNLKPESTD 480
 QY 703 FKADSEPEETPAPEVPEQVETEKVQAQLEAEVLLAKVTDSSLKANATETLAGLNNLT 762
 DB 481 TEETEEAEEDTTDEAEIPIQVENSVINAKIADAELEKVTDPISIRQNAMETLTGLKSL 540
 QY 763 LQIMDNNSIYAEKTLALLKSNPSSV 790
 DB 541 LGTKDNNTISAEDSLALLKESQAP 568

RESULT 48

AAB12728 ID AAB12728 standard; Protein; 334 AA.

XX AAB12728;

XX 21-NOV-2000 (first entry)

XX Streptococcus pneumoniae BVH-11A protein antigen SEQ ID NO:61.

XX Streptococcus pneumoniae; BVH-3; BVH-11; BVH-28; antigen; vaccine;

XX prophylaxis; therapy; infection; diagnosis; meningitis; bacteraemia;

XX otitis media; pneumonia; immunisation; bactericidal.

XX Streptococcus pneumoniae.

PN WO200039299-A2.
 XX 06-JUL-2000.
 XX 20-DEC-1999; 99WO-CA01218.
 XX 23-DEC-1998; 98US-0113800.
 XX (BIOC-) BIOCHEM PHARMA INC.
 PA Hamel J, Brodeur BR, Pineau I, Martin D, Rioux C, Charland N;
 XX WPI; 2000-452397/39.
 DR Streptococcal antigens useful for vaccinating against e.g. meningitis,
 PT otitis media, bacteraemia and/or pneumonia -
 CC Claim 18; Fig 26; 106pp; English.

The present invention describes nucleic acids (I) encoding protein
 CC antigens (II) from streptococcus pneumoniae. The protein antigens
 CC have bactericidal activity. The nucleic acids, encoding the protein
 CC antigens, may be used for the recombinant production of the proteins
 CC they encode. The protein antigens may then be used as vaccines for the
 CC prevention and treatment of streptococcal infections in mammals
 CC (especially humans) which result in, e.g. meningitis, otitis media,
 CC bacteraemia and/or pneumonia. The present sequence represents the
 CC *S. pneumoniae* BVH-11A protein antigen.

XX Sequence 334 AA;

Query Match 40.0%; Score 1666; DB 21; Length 334;
 Best Local Similarity 94.9%; Pred. No. 7.2e-115;
 Matches 316; Conservative 9; Mismatches 8; Indels 0; Gaps 0;

QY 1 SYELGLVQARTVKENNRVSYIDGKATQKTENLTTPDEVSKREGINAEQIVKITDQGYT 60
 DB 2 AVELGLHQAOQVKENNRVSYIDGKATQKTENLTTPDEVSKREGINAEQIVKITDQGYT 61
 QY 61 SHGHYHYNGKQVYDAIISBELMKDPNYKIKDEDIVNEVKGQVYKVGKYYVYLKDA 120
 DB 62 SHGHYHYNGKQVYDAIISBELMKDPNYKIKDEDIVNEVKGQVYKVGKYYVYLKDA 121
 QY 121 AHAENVRTKEEINQKQESHQREGTTPNDGAVALARSQGRYTTDGYIFNASDIEDT 180
 DB 122 AHAENVRTKEEINQKQESHQREGTTPNDGAVALARSQGRYTTDGYIFNASDIEDT 181
 QY 181 GDAYIVPHGDHYHYIPKNELSASELAEEAFLSGRNLNSRYTRRONSNTSRTNVFS 240
 DB 182 GDAYIVPHGDHYHYIPKNELSASELAEEAFLSGRNLNSRYTRRONSNTSRTNVFS 241
 QY 241 VSNPCTNTNTNSNNSNSQOSNDIDSLKQLYKPLSGRHHESDGLFPDPAQTSRT 300
 DB 242 VSNPCTNTNTNSNNSNSQOSNDIDSLKQLYKPLSGRHHESDGLFPDPAQTSRT 301
 QY 301 ARGVAVPHGDHYHYIPYQMSSELEERITARIPL 333
 DB 302 ARGVAVPHGDHYHYIPYQMSSELEERITARIPL 334

RESULT 49

AAU84027 ID AAU84027 standard; Peptide; 334 AA.

XX AAU84027;

XX 08-MAY-2002 (first entry)

XX Truncated variant of *S. pneumoniae* BVH-11, BVH-11A.

XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;

XX pneumonia; streptococcal bacterial infection; mutant; mutein.

```

OS Streptococcus pneumoniae.
OS Synthetic.
XX WO200198334-A2.
XX
XX 27-DEC-2001.
XX
XX 19-JUN-2001; 2001WO-CA00908.
XX
XX 20-JUN-2000; 2000US-212683P.
XX
XX (SHIR-) SHIRE BIOCHEM INC.
XX
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI; 2002-122272/16.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX epitope-bearing polypeptides, useful as vaccine components for treating
XX or preventing streptococcal infections such as otitis media,
XX meningitis, and bacteraemia.
XX
XX Example 1; Page -: 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
XX identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX comprising (I) is useful for therapeutic or prophylactic treatment of
XX meningitis, otitis media, bacteraemia or pneumonia infection in an
XX individual susceptible to these disorders. (II) is also useful for
XX therapeutic or prophylactic treatment of any streptococcal bacterial
XX infection (e.g., caused by Streptococcus pneumoniae, group A
XX Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
XX as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or
XX Streptococcus aureus) in an individual susceptible to the infection.
XX A polynucleotide (III) encoding (I) is useful in DNA immunisation
XX techniques. The Streptococcus polypeptides are useful in a diagnostic
XX test for S. pneumoniae infection. (III) is useful for designing DNA
XX probes for use in detecting the presence of Streptococcus in a biological
XX sample suspected of containing the bacteria. The DNA probes may also be
XX used for detecting circulating S. pneumonia nucleic acid in a sample for
XX diagnosing streptococcal infections. This sequence represents a truncate
XX of a Streptococcus pneumoniae gene used to obtain antigenic peptides.
XX Note: This sequence does not appear in the specification but has
XX been created according to information given in the invention.
XX
XX Sequence 334 AA;
XX
XX Query Match 40.0%; Score 1666; DB 23; Length 334;
XX Best Local Similarity 94.9%; Pred. No. 7.2e-115;
XX Matches 316; Conservative 9; Mismatches 8; Indels 0; Gaps 0;
XX
XX 1 SYELGIYQARTVKNRVSVIDKQATOKTENLTPEVSKREGINAEQIVIKITDGGYVT 60
XX 2 AVELGHAQOTVKNRVSVIDKQATOKTENLTPEVSKREGINAEQIVIKITDGGYVT 61
XX
XX 61 SHGDHYHYNGKPYPAIIISEELMDPNYKLDKEDIWNEVKGQYIKYDKGYVYLKQA 120
XX 62 SHGDHYHYNGKPYPAIIISEELMDPNYKLDKEDIWNEVKGQYIKYDKGYVYLKQA 121
XX
XX 121 AAADNVRTKEEIRKQEHSHQREGGTPRRDGAVALARSGRRTTDDGYIFNNSDIIEPT 180
XX 122 AAADNVRTKEEIRKQEHSHQREGGTSANDGVAARSGRRTTDDGYIFNNSDIIEPT 181
XX
XX 181 GDAYIVPHGDHYHYIPKNEISASBELAAEAFLSGRNLNSRTYRRQNSDNTSRINWVS 240
XX 182 GDAYIVPHGDHYHYIPKNEISASBELAAEAFLSGRNLNSRTYRRQNSDNTSRINWVS 241
XX
XX 241 VSNPQTNTNTSNTSNTSNGASQSNIDSLKQLYKLPISQRVESDGLVFDPAQITTSRT 300
XX 242 VSNPQTNTNTSNTSNTSNGASQSNIDSLKQLYKLPISQRVESDGLVFDPAQITTSRT 301

```

```

QY 301 ARGVAVPHGDHYHYIPYQSMSELEERIANRIPL 333
Db 302 ARGVAVPHGDHYHYIPYQSMSELEERIANRIPL 334
XX
XX RESULT 50
XX AAU84056
XX ID AAU84056 standard; Peptide; 1238 AA.
XX
XX AAU84056;
XX
XX 08-MAY-2002 (first entry)
XX
XX S. pneumoniae derived chimeric peptide, NEW29.
XX
XX BVH-3; BVH-11; vaccine; meningitis; otitis media; bacteraemia;
XX pneumonia; streptococcal bacterial infection; mutant; mutein;
XX BVH-11-2.
XX
XX Streptococcus pneumoniae.
XX Synthetic.
XX WO200198334-A2.
XX
XX 27-DEC-2001.
XX
XX 19-JUN-2001; 2001WO-CA00908.
XX
XX 20-JUN-2000; 2000US-212683P.
XX
XX (SHIR-) SHIRE BIOCHEM INC.
XX
XX Hamel J, Ouellet C, Charland N, Martin D, Brodeur B;
XX WPI; 2002-122272/16.
XX
XX New Streptococcus pneumoniae BVH-3 and BVH-11 variant and
XX epitope-bearing polypeptides, useful as vaccine components for treating
XX or preventing streptococcal infections such as otitis media,
XX meningitis, and bacteraemia.
XX
XX Example 1; Page -: 113pp; English.
XX
XX The invention describes an isolated polypeptide (I) with 70-90%
XX identity to Streptococcus pneumonia protein BVH-3, BVH-11, variants of
XX BVH-3 or BVH-11, or chimeric sequences derived from them. A vaccine (II)
XX comprising (I) is useful for therapeutic or prophylactic treatment of
XX meningitis, otitis media, bacteraemia or pneumonia infection in an
XX individual susceptible to these disorders. (II) is also useful for
XX therapeutic or prophylactic treatment of any streptococcal bacterial
XX infection (e.g., caused by Streptococcus pneumoniae, group A
XX Streptococcus such as Streptococcus pyogenes, group B Streptococcus such
XX as Streptococcus agalactiae, S. dysgalactiae, S. uberis, S. norcardia or
XX Streptococcus aureus) in an individual susceptible to the infection.
XX A polynucleotide (III) encoding (I) is useful in DNA immunisation
XX techniques. The Streptococcus polypeptides are useful in a diagnostic
XX test for S. pneumoniae infection. (III) is useful for designing DNA
XX probes for use in detecting the presence of Streptococcus in a biological
XX sample suspected of containing the bacteria. The DNA probes may also be
XX used for detecting circulating S. pneumonia nucleic acid in a sample for
XX diagnosing streptococcal infections. This sequence represents a chimeric
XX gene created from fragments and variant fragments of Streptococcus
XX pneumoniae genes, described in the method of the invention.
XX Note: This sequence does not appear in the specification but has
XX been created according to information given in the invention.
XX
XX Sequence 1238 AA;
XX
XX Query Match 39.9%; Score 1661; DB 23; Length 1238;
XX Best Local Similarity 66.5%; Pred. No. 1.3e-113;
XX Matches 314; Conservative 57; Mismatches 97; Indels 4; Gaps 2;
XX
XX 266 DIDSLKQLYKLPISQRVESDGLVFDPAQITTSRTARGVAVPHGDHYHYIPYQSMSELEE 325

```

```

Db      2 DIDSLKQYLKPLSQRHVESDGLIFDPAQISRTARGAVPHGNHHTIPEQMSLEK 61
QY      326 RIARIIPLRYSNHWPPDSRPEQSPQPTPEPSPGQPAANLK-IDSN---SSLVSQLYR 381
Db      62 RIARIIPLRYSNHWPPDSRPEPSPQPTPEPSPGQPAANPQAPASNPIDKLVKEAVR 121
QY      382 KYGEGYVFEKGISRYVFAKDLPSFTVKNLSEKLSKQESVSHTLTAKKENVAPRDOFPYD 441
Db      122 KYGDSYVEENCVSRIIPAKNLSAETAGIDSKLAKQESLSHKLQAKKTDLPPSSDRFYN 181
QY      442 KAYNLTEAHKALFYNKGKNSDQALDKLIERLNDSTNKEKLVDDLAFAPITHPERL 501
Db      182 KAYDLARIHQDLNKGQVDFEALDNLRLKDVSSDKVCLVDDILAFLAPIRHPERL 241
QY      502 GKPNQIETEBEVRIAQLADKYTTSBGYIFDEHDIISDEGDAYVTPHKGSHWIGKDSL 561
Db      242 GKPNQIETEBEVRIAQLADKYTTSBGYIFDEHDIISDEGDAYVTPHKGSHWIGKDSL 301
QY      562 SDKEKVAQAAYTKEKGIILPPSPADVKANPTGSAATYNRVKEKRIPLVRLPYMVEHT 621
Db      302 SEAEPAQAAYTKEKGIILPPSPADVKANPTGSAATYNRVKEKRIPLVRLPYMVEHT 361
QY      622 VEKNGNLIIPKDHYNHNIKFAFPDDHTYKAPNGYTLLEDLPATIKYVEHPDERPHSNDG 681
Db      362 VEKNGNLIIPKDHYNHNIKFAFPDDHTYKAPNGYTLLEDLPATIKYVEHPDERPHSNDG 421
QY      682 WGNASEHVLGKKHSEDPNKNFKADEBVEEETPAEBEVQVETEKVEAOLKE 733
Db      422 FGNASDHVGNMOPSQLSYSTASDNNTOSVAKGSTSKPANKSENLOSILKE 473

```

Search completed: November 14, 2003, 10:34:12
 Job time : 89 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 10:32:55 / Search time 41 Seconds
(without alignments)
1867.080 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 4165

Sequence: 1 SYELGLYQARTVKNRVS.....KLIALIKGSPSSVSKKIN 796

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : PIR_76:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4163	100.0	802	2	C95136	conserved domain p
2	4159	99.9	828	2	E98004	hypothetical prote
3	3224	77.4	855	2	D98004	histidine Motif-Co
4	2797.5	67.2	819	2	B95136	conserved domain p
5	2772	66.6	839	2	G95115	conserved hypotet
6	2733	65.6	853	2	C97985	hypothetical prote
7	1246	29.9	1039	2	H95115	conserved hypotet
8	1243	29.8	1039	2	D97985	hypothetical prote
9	934	22.4	822	2	T46758	hypothetical prote
10	243	5.8	182	2	F97985	hypothetical prote
11	192.5	4.6	1390	2	T14004	trifla protein - sli
12	176.5	4.2	1271	2	A45555	glutamate rich pro
13	172	4.1	2481	2	D90011	FmtB protein (impo
14	171	4.1	1233	2	S56271	hypothetical prote
15	168.5	4.0	1043	2	D84900	hypothetical prote
16	165	3.9	891	2	G89957	hypothetical prote
17	161	3.9	1420	1	A44361	amiloride-sensitiv
18	159	3.8	2810	2	T22298	hypothetical prote
19	158	3.8	1650	2	T18444	hypothetical prote
20	157	3.8	910	1	S73361	dnau homolog prote
21	157	3.8	1658	2	S55103	hypothetical prote
22	157	3.8	2195	2	S61103	SEC16 protein - ye
23	156.5	3.8	1495	2	T48429	hypothetical prote
24	156	3.7	519	2	G84598	probable bZIP tran
25	154.5	3.7	1338	2	T18416	hypothetical prote
26	154.5	3.7	1463	2	T30290	AAS surface protei
27	154.5	3.7	1516	2	E71619	RAD2 endonuclease
28	154.5	3.7	3924	2	S37431	ankyrin 2, neurona
29	154	3.7	749	2	S23467	probable long-chain

30	151.5	3.6	1163	2	A36685	205K microtubule-a
31	150.5	3.6	1176	2	S66771	hypothetical prote
32	150	3.6	1132	2	B82538	ribonuclease E Xf2
33	150	3.6	2401	2	T28676	rhopty protein
34	149.5	3.6	1175	2	S51005	protein-tyrosine-p
35	149	3.6	1274	2	A89959	hypothetical prote
36	149	3.6	2829	2	A42771	reticulocyte-bind
37	149	3.6	4688	2	F82885	hypothetical prote
38	149	3.6	6713	2	B89921	hypothetical prote
39	148.5	3.6	1785	2	T21558	hypothetical prote
40	148.5	3.6	1959	2	AG1085	hypothetical prote
41	148.5	3.6	1969	2	T38495	hypothetical prote
42	148	3.6	980	2	E71606	hypothetical prote
43	148	3.6	1230	2	T22458	hypothetical prote
44	148	3.6	1856	2	C95008	immunoglobulin A1
45	147	3.5	700	2	A56976	transfer complex p
46	147	3.5	1269	2	F84730	probable myosin he
47	147	3.5	1929	2	T21559	hypothetical prote
48	146.5	3.5	2004	2	F95133	immunoglobulin A1
49	146	3.5	821	2	S67087	hypothetical prote
50	145.5	3.5	1073	2	S14032	kinasin-related pr
51	145.5	3.5	1085	2	T38378	kinasin-like prote
52	145.5	3.5	1240	2	S52734	hypothetical prote
53	144.5	3.5	934	2	T47546	protein kinase-lik
54	144.5	3.5	4152	2	T31102	filamentous hemag
55	144.5	3.5	4919	2	T31105	hypothetical prote
56	143.5	3.4	900	2	C64232	alanine-tRNA ligas
57	142.5	3.4	875	2	S70115	ZIP1 protein - yea
58	142.5	3.4	1066	2	T45283	growth polarity ma
59	142.5	3.4	1066	2	T41099	staurosporine targ
60	142.5	3.4	1979	2	T71622	hypothetical prote
61	142.5	3.4	5170	2	T15348	hypothetical prote
62	142	3.4	1257	2	T00486	serine/threonine-s
63	142	3.4	1744	2	UH0720	Canabin - African
64	142	3.4	1822	2	S33441	EF protein - Strep
65	141.5	3.4	490	2	S52830	HMS1 protein - yea
66	141.5	3.4	1315	2	T28679	fibronogen-binding
67	141.5	3.4	1928	2	S46773	myosin heavy chain
68	141.5	3.4	2214	1	A46151	protein-tyrosine-p
69	141.5	3.4	2748	2	S57976	nuclear migration
70	141.5	3.4	3498	2	T22330	hypothetical prote
71	141	3.4	914	2	B48086	translation initia
72	141	3.4	1190	2	T47536	oxysterol-binding
73	141	3.4	1280	2	T42514	kinase anchor prot
74	141	3.4	1332	2	S41552	probable transcrip
75	141	3.4	2288	2	T29999	hypothetical prote
76	141	3.4	3890	2	C89921	hypothetical prote
77	140.5	3.4	507	2	S05542	hypothetical prote
78	140.5	3.4	719	2	A81358	hypothetical prote
79	140.5	3.4	2231	2	T49042	hypothetical prote
80	139.5	3.3	1535	2	D71870	hypothetical prote
81	139.5	3.3	1609	1	MMHUB2	laminin gamma-1 ch
82	139.5	3.3	4549	2	T20771	hypothetical prote
83	139.5	3.3	4667	2	T20774	hypothetical prote
84	139	3.3	1247	2	C89583	protein K07E3.1 (i
85	139	3.3	1385	2	D89824	hypothetical prote
86	139	3.3	1957	2	T38077	hypothetical prote
87	138.5	3.3	1703	2	S15047	SNF2 protein - yea
88	138.5	3.3	2269	2	T28677	rhopty protein -
89	138	3.3	1639	2	S05603	major mercozite su
90	138	3.3	2295	2	B71621	probable membrane
91	137.5	3.3	1099	2	G90546	conserved hypotet
92	137.5	3.3	1510	2	T33100	hypothetical prote
93	137.5	3.3	1714	2	E71609	Ser/Thr protein ki
94	137.5	3.3	1790	2	S67593	transport protein
95	137.5	3.3	2175	1	S03170	homeotic protein c
96	137.5	3.3	2364	2	A56577	microtubule-associ
97	137.5	3.3	2484	2	T26216	hypothetical prote
98	137.5	3.3	2607	2	T26215	hypothetical prote
99	137	3.3	990	2	H86293	protein T24D18.4 (
100	137	3.3	1032	2	T30270	major tegumental a

ALIGNMENTS

RESULT 1

C95136

conserved domain protein SP1175 [imported] - Streptococcus pneumoniae (strain TIGR4)

C:Species: Streptococcus pneumoniae

C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001

C:Accession: C95136

R:Retcelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heidon, J.D.; Unayam, L.A.; White, O.; Salsbery, S.L.; Lewis, M.R.; Radune, D.; Holtzaple, nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001

A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,

A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.

A:Reference number: A95000; MUID:21357209; PMID:11463916

A:Accession: C95136

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-802 <KUR>

A:Cross-references: GB:AE005672; PIDN:AAK75284.1; PID:g14972655; GSPDB:GN00164; TIGR:SP4

A:Experimental source: strain TIGR4

C:Genetics:

C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 100.0%; Score 4163; DB 2; Length 802;

Best Local Similarity 99.9%; Pred. No. 1.7e-219;

Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy	1	SYELGLVQARTVKENNRVS	YIDGKQATQKTEMLTPDEVSKREGINAEOIVIKITDQGYVT	60
Db	7	SYELGLVQARTVKENNRVS	YIDGKQATQKTEMLTPDEVSKREGINAEOIVIKITDQGYVT	66
Qy	61	SHGDHYHYNGKVPYDAI	ISELLMKDPNYKLKDEIVNEVKGVIKVDGKYVYLLKDA	120
Db	67	SHGDHYHYNGKVPYDAI	ISELLMKDPNYKLKDEIVNEVKGVIKVDGKYVYLLKDA	126
Qy	121	AHADNVRKEEINRQKQES	QHRREGGTPRNDGAVALLASQGYTTDDGYIFNADIIEDT	180
Db	127	AHADNVRKEEINRQKQES	QHRREGGTPRNDGAVALLASQGYTTDDGYIFNADIIEDT	186
Qy	181	GDAYIVPHGDHYHYIPK	NELASSELAAEAFLSGRGLNSRTRYRQNSDTSRTNWPS	240
Db	187	GDAYIVPHGDHYHYIPK	NELASSELAAEAFLSGRGLNSRTRYRQNSDTSRTNWPS	246
Qy	241	VSNPGTTNTNTSNNST	NSQASQNSNDISLLKQLYKPLSGRHVESDGLVDPDAQITSRT	300
Db	247	VSNPGTTNTNTSNNST	NSQASQNSNDISLLKQLYKPLSGRHVESDGLVDPDAQITSRT	306
Qy	301	ARGVAVPHGDHYHYIP	YSQMSLEERIRIIPLRYSNHWVPDSRPEQSPQPTPEPSPG	360
Db	307	ARGVAVPHGDHYHYIP	YSQMSLEERIRIIPLRYSNHWVPDSRPEQSPQPTPEPSPG	366
Qy	361	POPAPNLKIDSSNSL	SVQLVKRVGEGYVFEKGISRYVPAKDLPEETVKNLESLSKQES	420
Db	367	POPAPNLKIDSSNSL	SVQLVKRVGEGYVFEKGISRYVPAKDLPEETVKNLESLSKQES	426
Qy	421	VSHLTAKENAVAPRDO	EFDYKAYNLLTEAHKALFXNGRNSDFQALDKLLERLNDSTN	480
Db	427	VSHLTAKENAVAPRDO	EFDYKAYNLLTEAHKALFXNGRNSDFQALDKLLERLNDSTN	486
Qy	481	KEKLVDDLAFAPITHT	PERLGKPNISOETEDEVRIQLADKYTTSQGYIFDEHDIISD	540
Db	487	KEKLVDDLAFAPITHT	PERLGKPNISOETEDEVRIQLADKYTTSQGYIFDEHDIISD	546
Qy	541	EGDAVTPHMGSHHIG	KDSLSDKEKVAQAATYTKGILPSPDADVANKPTGDSAAIY	600
Db	547	EGDAVTPHMGSHHIG	KDSLSDKEKVAQAATYTKGILPSPDADVANKPTGDSAAIY	606
Qy	601	NRVKGKKIPLVRLPE	VYWEHTVEVKNGLIIPKHQHNKIKFAMPDDHTTYKAPNGYTLSD	660
Db	607	NRVKGKKIPLVRLPE	VYWEHTVEVKNGLIIPKHQHNKIKFAMPDDHTTYKAPNGYTLSD	666

RESULT 2

E98004

hypothetical protein phca [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001

C:Accession: E98004

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; Ee, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: E98004

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-828 <KUR>

A:Cross-references: GB:AE007317; PIDN:AAK99865.1; PID:g15458683; GSPDB:GN00174

C:Genetics:

C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 99.9%; Score 4159; DB 2; Length 828;

Best Local Similarity 99.7%; Pred. No. 3e-219;

Matches 794; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy	1	SYELGLVQARTVKENNRVS	YIDGKQATQKTEMLTPDEVSKREGINAEOIVIKITDQGYVT	60
Db	33	SYELGLVQARTVKENNRVS	YIDGKQATQKTEMLTPDEVSKREGINAEOIVIKITDQGYVT	92
Qy	61	SHGDHYHYNGKVPYDAI	ISELLMKDPNYKLKDEIVNEVKGVIKVDGKYVYLLKDA	120
Db	93	SHGDHYHYNGKVPYDAI	ISELLMKDPNYKLKDEIVNEVKGVIKVDGKYVYLLKDA	152
Qy	121	AHADNVRKEEINRQKQES	QHRREGGTPRNDGAVALLASQGYTTDDGYIFNADIIEDT	180
Db	153	AHADNVRKEEINRQKQES	QHRREGGTPRNDGAVALLASQGYTTDDGYIFNADIIEDT	212
Qy	181	GDAYIVPHGDHYHYIPK	NELASSELAAEAFLSGRGLNSRTRYRQNSDTSRTNWPS	240
Db	213	GDAYIVPHGDHYHYIPK	NELASSELAAEAFLSGRGLNSRTRYRQNSDTSRTNWPS	272
Qy	241	VSNPGTTNTNTSNNST	NSQASQNSNDISLLKQLYKPLSGRHVESDGLVDPDAQITSRT	300
Db	273	VSNPGTTNTNTSNNST	NSQASQNSNDISLLKQLYKPLSGRHVESDGLVDPDAQITSRT	332
Qy	301	ARGVAVPHGDHYHYIP	YSQMSLEERIRIIPLRYSNHWVPDSRPEQSPQPTPEPSPG	360
Db	333	ARGVAVPHGDHYHYIP	YSQMSLEERIRIIPLRYSNHWVPDSRPEQSPQPTPEPSPG	392
Qy	361	POPAPNLKIDSSNSL	SVQLVKRVGEGYVFEKGISRYVPAKDLPEETVKNLESLSKQES	420
Db	393	POPAPNLKIDSSNSL	SVQLVKRVGEGYVFEKGISRYVPAKDLPEETVKNLESLSKQES	452
Qy	421	VSHLTAKENAVAPRDO	EFDYKAYNLLTEAHKALFXNGRNSDFQALDKLLERLNDSTN	480
Db	453	VSHLTAKENAVAPRDO	EFDYKAYNLLTEAHKALFXNGRNSDFQALDKLLERLNDSTN	512
Qy	481	KEKLVDDLAFAPITHT	PERLGKPNISOETEDEVRIQLADKYTTSQGYIFDEHDIISD	540

```

Db      513 KEELVDLLAFLAPLPIHPERLGRKPNQIETEDBEVIAQLADKYTTSDGIYFDEHDIISD 572
QY      541 EGDAYVTPHMGSHWIGKDSLDEKVAQAAYTKERGIIPSPDPAVKANPTGDSAAAY 600
Db      573 EGDAYVTPHMGSHWIGKDSLDEKVAQAAYTKERGIIPSPDPAVKANPTGDSAAAY 632
QY      601 NEVKGKRIPLVRLPYMVEHTVEVKNGLIIPKDHVHNIKFAMFDHTYKAPNGYTLSD 660
Db      633 NEVKGKRIPLVRLPYMVEHTVEVKNGLIIPKDHVHNIKFAMFDHTYKAPNGYTLSD 692
QY      661 LEATIKYVYEHDPDRPHSDGNGNASEHVLGKKHSEDNPKNKADEEPEETPAPEVP 720
Db      693 LEATIKYVYEHDPDRPHSDGNGNASEHVLGKKHSEDNPKNKADEEPEETPAPEVP 752
QY      721 QVETEKVEAQLKEAEVLAKVTDLSLKANAETLTLQIMDNNSIYMAEAKLTA 780
Db      753 QVETEKVEAQLKEAEVLAKVTDLSLKANAETLTLQIMDNNSIYMAEAKLTA 812
QY      781 LKGSNPPSSVSKKIN 796
Db      813 LKGSNPPSSVSKKIN 828

```

RESULT 3

```

D96004 histidine Motif-Containing protein [imported] - Streptococcus pneumoniae (strain R6)
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: D96004
R:Host: J.A.; Albom Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; Dehoff, B.S.; E
R, R.; Leblanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; Mchren, S.; M
Y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: D96004
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-855 <KUR>
A:Cross-references: GB:AE007317; PIDN:AAK99864.1; PID:g15458682; GSPDB:GN00174
C:Genetics:
A:Gene: phpA
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

```

```

Query Match      77.4%; Score 3224; DB 2; Length 855;
Best Local Similarity 75.0%; Pred. No. 3.2e-168;
Matches 617; Conservative 69; Mismatches 99; Indels 38; Gaps 4;

QY      1 STELGIYQARTYKENVRSYIDGKATOKTENTLTPDEVSKREGINAEQIVIKITDGGYV 60
Db      32 AVELGIAQYQYKENVRSYIDGKATOKTENTLTPDEVSKREGINAEQIVIKITDGGYV 91
QY      61 SHGDHYHYNGKVPYDAIISEELMKDPNYKLKEDIYNEVKGVIYVDGKYVYVYLKD 120
Db      92 SHGDHYHYNGKVPYDAIISEELMKDPNYQLDEIYISIKGVYIKDGVYVYVYLKD 151
QY      121 AAADNVRTKEINRQKQSHQREHGTFRNDGVALARSQGRYTTDDGYIFNADIIEDT 180
Db      152 AAADNVRTKEINRQKQSHQREHGTFRNDGVALARSQGRYTTDDGYIFNADIIEDT 211
QY      181 GDAYIVPHGDHYHYIKNELSASELAAEAFLSGRNLNSRTYRQNSDNTSRTTWVPS 240
Db      212 GDAYIVPHGDHYHYIKNELSASELAAEAFLSGRNLNSRTYRQNSDNTSRTTWVPS 271
QY      241 VSNPGTTNTSNTSNTSNTSNTSNTSNTSNTSNTSNTSNTSNTSNTSNTSNTSNTS 300
Db      272 VSNPGTTNTSNTSNTSNTSNTSNTSNTSNTSNTSNTSNTSNTSNTSNTSNTSNTS 331
QY      301 AGVAVPHGDHYHYIFYSQMSLEERIAIIRLYSNHNVDPDRPEQSPQPTPEPSPQ 360
Db      332 AGVAVPHGDHYHYIFYSQMSLEERIAIIRLYSNHNVDPDRPEQSPQPTPEPSPS 391

```

```

QY      361 POPAPNLK-IDSN---SSIVSQLVRKVGEGYFEKGISRYVFAKDLPSETVKNLESKLS 416
Db      392 POPAPNPQAPNPPIDEKLVKAIVKVGDIYFEENGVRITPAKDLSETAAGIDSKLA 451
QY      417 KOESVSHLTAKENAVAPRQGFYDKAYNLLTEAKALEXKNGRNSDFQALDKLERND 476
Db      452 KOESVSHLTAKENAVAPRQGFYDKAYNLLTEAKALEXKNGRNSDFQALDKLERND 511
QY      477 ESTNKEVYDDLAFAPLPIHPERLGRKPNQIETEDBEVIAQLADKYTTSDGIYFDEH 536
Db      512 VSSDKVAVDDILAFAPLPIHPERLGRKPNQIETEDBEVIAQLADKYTTSDGIYFDEH 571
QY      537 IISDGDVAVTPHMGSHWIGKDSLDEKVAQAAYTKERGIIPSPDPAVKANPTGDSA 596
Db      572 IISDGDVAVTPHMGSHWIGKDSLDEKVAQAAYTKERGIIPSPDPAVKANPTGDSA 631
QY      597 AAIYVRVKGERIPLVRLPYMVEHTVEVKNGLIIPKDHVHNIKFAMFDHTYKAPNGY 656
Db      632 AAIYVRVKGERIPLVRLPYMVEHTVEVKNGLIIPKDHVHNIKFAMFDHTYKAPNGY 691
QY      657 TLEDLPATIKYVYEHDPDRPHSDGNGNASEHVLGKKHSEDNPKNKADEEPEETPA 710
Db      692 TLEDLPATIKYVYEHDPDRPHSDGNGNASEHVLGKKHSEDNPKNKADEEPEETPA 751
QY      711 EET-----PAPEVQVETEKVEAQLKEAEVLAKV 742
Db      752 EETPEBEKQKSPSPKPTPEEPEEPEEPEEPEEPEEPEEPEEPEEPEEPEEPEEPEE 811
QY      743 DSSLKANATETLAGLRNNLTQIMDNNSIYMAEAKLTLKGS 785
Db      812 NIIISNKKETITGLKNNLTFTQDNNTIYMAEAKLTLKGS 854

```

RESULT 4

```

B95136 conserved domain protein SP1174 [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: B95136
R:Retelien, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Hei
om, J.D.; Umayam, L.A.; White, O.; Salasberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfel,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.D.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: B95136
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-819 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75283.1; PID:g14972654; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP1174
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

```

```

Query Match      67.2%; Score 2797.5; DB 2; Length 819;
Best Local Similarity 67.0%; Pred. No. 5.7e-145;
Matches 546; Conservative 84; Mismatches 138; Indels 47; Gaps 7;

QY      1 STELGIYQAR-KTVKENVRSYIDGKATOKTENTLTPDEVSKREGINAEQIVIKITDGGYV 59
Db      21 STELGIYQARQKQKSSNRVAYIDGQAKANLTPDEVSKREGINAEQIVIKITDGGYV 80
QY      60 TSHGDHYHYNGKVPYDAIISEELMKDPNYKLKEDIYNEVKGVIYVDGKYVYVYLKD 119
Db      81 TSHGDHYHYNGKVPYDAIISEELMKDPNYQLKSDIYNEIKGVYIKDGVYVYVYLKD 140
QY      120 AAADNVRTKEINRQKQSHQREHGTFRNDGVALARSQGRYTTDDGYIFNADIIED 179
Db      141 AAADNVRTKEINRQKQSHQREHGTFRNDGVALARSQGRYTTDDGYIFNADIIED 197

```

```

Oy 180 TGDVAVPHGDHYHYIIPKNEISASELAAEAFLSGRGNLSNRTYRQNSDTSRTNWP 239
Db 198 TGDVAVPHGDHYHYIIPKNEISASELAAEAFLSGRGNLSNRTYRQNSDTSRTNWP 246
Oy 240 SVSNNGTTNTNTSNNSTNSQASQNSDIDSLKQLYKPLSGRHYESDGLVDPQITSR 299
Db 247 NPAQRLSENHNLVTPTPHQ--NOGENISSILRELYAKPLSERHVESDGLTFDPAQITSR 305
Oy 300 TARGVAVPHGDHYHYIIPYQOMSELEERARIIPLYRSNHWVPSRPEQSPQPTPEPSP 359
Db 306 TARGVAVPHGDHYHYIIPYQOMSELEERARIIPLYRSNHWVPSRPEQSPQPTPEPSP 365
Oy 360 GPAPAPNLKIDSNSSIVSQLVKRGVGEVFEKGISRYVPAKDLPSFTVKNLESKLKSGOE 419
Db 366 SPQAPNPQAPSNPID--EKLVEAVKAVKVDGYFEENGVSRYIPAKDLSETAAGIDSLAAGE 423
Oy 420 SVSHTLTAKKENAVAPRDOEFYDKAVNLTLEAKALFXNKGNSDFQALDKLERLNDEST 479
Db 424 SLSHKLTGTRKTDLPSSDREFFYNKAYDLARIHQDLDNKGRQVDFEALDNLERLKDVS 483
Oy 480 NKEKLVDDLAFIAPITPERLGRKNSQIEYTEDEVRIAOADKYTSDGYIFDEHDITIS 539
Db 484 DKVKVEBILAFIAPITPERLGRKNAQIITYTDEIOVAKLAGKTYTDEGYIFDPRDITS 543
Oy 540 DEGDVAVTPHMGSHWIKDLSDEKEXAAQAAYTEKGLPPSPDADYKANPTGDSAAAI 599
Db 544 DEGDVAVTPHMGSHWIKDLSDEKEXAAQAAYTEKGLTPSPDADYKANPTGDSAAAI 603
Oy 600 YNRVAKGKRIPLVRLPYVEHTVEVNGNLIIPHKHYNHIKFAFPDHTTYKAPNGYLE 659
Db 604 YNRVAKKAVPLDRMPVNLQYTVENVNGSLIIPHYDHYNHIKFEWFDGLYEAPKGYTLE 663
Oy 660 DLFAITKYVVEHPPDERPHSNDGWNASEHVLGKHSDSDPKNFADE-----PVET 713
Db 664 DLATVATYVEHPPDERPHSNDGWNASEHVLGKHSDSDPKNFADE-----PVET 723
Oy 714 -----PAPEVPOVETEKVEAQLKEAEVLAKYTDLSLKANA 750
Db 724 PREKPOSEKPEPRKPTPEPESPESEPEVETEKVEKLEAEDLLGKIQDPIIKNA 783
Oy 751 TETLAGLRNNLTQIMDNNSIWAEEKLLALIKGS 785
Db 784 KETLTGLKNNLFTQDNNTIWAEEKLLALIKES 818

RESULT 5
G95115
C:Species: hypothetical protein SP1003 [imported] - Streptococcus pneumoniae (strain TIGR
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 24-Aug-2001
C:Accession: G95115
R:Tetrelain, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: G95115
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-839 <KUR>
A:Cross-references: GB:AF005672; PIDN:AAK5120.1; PID:g14972476; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
C:Superfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match 66.6%; Score 2772; DB 2: Length 839;
Best Local Similarity 65.0%; Pred. No. 1.4e-143;
Matches 541; Conservative 91; Mismatches 144; Indels 56; Gaps 8;

Oy 1 SYELGLVQARTV-KENNRVSYIDGQATQKNTENLTPEVSKREGINAEQIVIKITDQGV 59

```

```

Db 21 SYELGRHOAGVKKESNRVSYIDGQAGKXENLTPEVSKREGINAEQIVIKITDQGV 80
Oy 60 TSHGDHYHYNGKPYPAIIISEELMKDPNKKLDEQIVNEKGVYIKVNGKYVVLKD 119
Db 81 TSHGDHYHYNGKPYPAIIISEELMKDPNKKLDEQIVNEKGVYIKVNGKYVVLKD 140
Oy 120 AAHADNRTKEIRKQKESQREGTTPNDGVAALARSQGRYTTDGYIFNADITIED 179
Db 141 AAHADNRTKEIRKQKESQREGTTPNDGVAALARSQGRYTTDGYIFNADITIED 198
Oy 180 TGDVAVPHGDHYHYIIPKNEISASELAAEAFLSGRGNLSNRTYRQNSDTSRTNWP 239
Db 199 TGDVAVPHGDHYHYIIPKNEISASELAAEAFLSGRGNLSNRTYRQNSDTSRTNWP 247
Oy 240 SVSNNGTTNTNTSNNSTNSQASQNSDIDSLKQLYKPLSGRHYESDGLVDPQITSR 299
Db 248 NPAQRLSENHNLVTPTPHQ--NOGENISSILRELYAKPLSERHVESDGLTFDPAQITSR 306
Oy 300 TARGVAVPHGDHYHYIIPYQOMSELEERARIIPLYRSNHWVPSRPEQSPQPTPEPSP 359
Db 307 TARGVAVPHGDHYHYIIPYQOMSELEERARIIPLYRSNHWVPSRPEQSPQPTPEPSP 366
Oy 360 GPAPAPNLK-IDSN--SSIVSQLVKRGVGEVFEKGISRYVPAKDLPSFTVKNLESKL 415
Db 367 SPQAPNPQAPSNPID-EKLVEAVKAVKVDGYFEENGVSRYIPAKDLSETAAGIDSLK 426
Oy 416 SKQSSVSHTLTAKKENAVAPRDOEFYDKAVNLTLEAKALFXNKGNSDFQALDKLERLN 475
Db 427 AKQSSLSHKLGAKKTDLPSSDREFFYNKAYDLARIHQDLDNKGRQVDFEALDNLERL 486
Oy 476 DESTNKEKLVDDLAFIAPITPERLGRKNSQIEYTEDEVRIAOADKYTSDGYIFDEH 535
Db 487 DVPSDKYKLVNDILAFIAPITPERLGRKNAQIITYTDEIOVAKLAGKTYTDEGYIFDPR 546
Oy 536 DIISDEGDVAVTPHMGSHWIKDLSDEKEXAAQAAYTEKGLPPSPDADYKANPTGDS 595
Db 547 DIISDEGDVAVTPHMGSHWIKDLSDEKEXAAQAAYTEKGLTPSPDADYKANPTGDS 606
Oy 596 AALAYNRVAKGKRIPLVRLPYVEHTVEVNGNLIIPHKHYNHIKFAFPDHTTYKAPNG 655
Db 607 AALAYNRVAKGKRIPLVRLPYVEHTVEVNGNLIIPHKHYNHIKFAFPDHTTYKAPNG 666
Oy 656 YTTEDLFAITKYVVEHPPDERPHSNDGWNASEHVLGK-----KDHS----- 697
Db 667 YTTEDLFAITKYVVEHPPDERPHSNDGWNASEHVLGK-----KDHS----- 726
Oy 698 -----DPNKNFKADEPVEETPAPEVPOVETEKVEAQLKEAEVL 738
Db 727 ESDKENHAGLPSADNLYKPSDTPEEAEEDTDEAEIPQVENSVINAKIADAEL 786
Oy 739 AKYTSSLSKANATETLAGLRNNLTQIMDNNSIWAEEKLLALIKGNSPSSV 790
Db 787 EKVTDPSIRONAMETLTGLKSSLTGKDNNTISAEVDSLALIKESQAPAI 838

RESULT 6
G97985
C:Species: hypothetical protein phd [imported] - Streptococcus pneumoniae (strain R6)
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 02-Nov-2001
C:Accession: G97985
R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Buggett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, U.N.; Lefkowitz, B.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544234
A:Accession: G97985
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-853 <KUR>

```

A:Cross-references: GB:AE007337; PIDN:AAK9711.1; PID:GL5458514; GSEDB:GN00174
C:Genetics:
A:Gene: phd
C:Superfamily: Streptococcus agalactiae hypothetical 92.4k protein

Db 612 QNF 614

RESULT 8
D97985
hypochemical protein phle [imported] - Streptococcus pneumoniae (strain R6)
C|Species: Streptococcus pneumoniae
C|Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
A|Accession: D97985
R|Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E
e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, R.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A|Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A|Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A|Reference number: A97872; MUID:21429245; PMID:11544234
A|Accession: D97985
A|Status: Preliminary
A|Molecule type: DNA
A|Residues: 1-1039 <KUR>
A|Cross-references: GB:AE007317; PIDN:AAK99712.1; PID:q15458515; GSPDB:GN00174
C|Genetics:
A|Gene: phle

Query Match	29.8%;	Score 1243;	DB 2;	Length 1039;
Best Local Similarity	41.2%;	Pred. No. 5.2e-60;		
Matches 298;	Conservative 94;	Mismatches 180;	Indels 152;	Gaps 20

```

OY 1 YEELGLQARIVYKNNRNVSYIDGKOATQKTEMLTPDEKSKREGINAEDIVKITDQGVYT 60
Db 22 AYALQOHSSQENKDNKNNRVSYVDGSSQSQISEMLTPDOVSQKEGJOAEQIVKITDQGVYT 81
OY 61 SHGDHYHYNGKVPYDAIISEBLLMKDPNYKLKEDIVNEYKGGYVIRVDKRYVYVYLKDA 120
Db 82 SHGDHYHYNGKVPYDALFSEELLMKDPNYQKLDADIVNEYKGGYIIRVDKRYVYVYLKDA 141
OY 121 AHADNVRIRKEELNROKQESHQREGCTPRNDCAVALASQGRYTTDDGIVFNASIIEDT 180
Db 142 AHADNVRIRKDEINROKQEHVKNONE---KVNSNVAVASQGRYTTNDGYVFNPAIIIEDT 197
OY 181 GDAYVVPBGDHYHIFPKNELSASELAAAEAFISGCGNLSNSRTYRQSDNTRTWNVPS 240
Db 198 GNAYVVPBGCHYHYIFKSDLSASELAAAKAHLAGNMQPSQLSYSTSTASDN----- 248
OY 241 VSNPCTNTNTNSNNTNSQASQSDIDSLKQLYKPLSQPHVSDGLVPDPAQITRT 300
Db 249 -----NTOSVAKGSTSKPANKSENLOSLLKELTSDSPSAORYSDGLVFPDPAKIIISRT 301
OY 301 ARGVAVPHGDHYHIFPYGOMSELEERIRARIIPLRYSNHWWPDDSRPBDSPQRPPEPSPG 360
Db 302 PNGVAIIPBGDHYHIFPYSKLSALEEKIARKVP----- 333
OY 361 POPAPNLKIDSNSLSIVOLVRKVGEGYVFEKEGISRYVFAKDLPSBETVKNLESKLSKQES 420
Db 334 -----ISGHTSVSTINAK-----PNEVVSSIGSLSNSPSS 363
OY 421 VSHITLAKKENVAPRPDOEFYDKANYLLPTEAHKALFXKNGRNSDPALDKLERLNDESTN 480
Db 364 ---LTSKEISSASDGYIFNPK-DIVBETATAYIVRRG-DHFYIRSK-----SNOIG 410
OY 481 KEKLVDDLLAFLAPITHEBERLQKP-NSQIEYEYDEVRITAOQLADKYTSDGYIFDEHDIIS 539
Db 411 QPTLPNNSLATPSP-----SLPFINPGISHEKHE-----EDGIGFQDNRRIIA 451
OY 540 DEGDAYVTPPHMGHSWIGKDSLSDEKEVAAQAYTEKEGILPPSPDADYKANPTG-DSAAA 598
Db 452 EDESGFISHNGHNHYFFPKKDLTEBOIKAAQKHLEE-----YKTSHNGIDSLSS 500
OY 599 IYNVRKSG-----EKRIPLVRLPYWV---EHTVEVNGNLIIPHHDIYNIKFA 643
Db 501 HEODYPGNAKEMKMDKKEIKIAGIMKOYGVYKESIVYVNEKEKAIITPPHGDHHAADP- 558
OY 644 WFDHDTYKAPNGYLTLEDLFAITIKYVVEHPDERPHSNDG---WGNASEHVLGKKDHSBDP 699

```

```

Db      559  -IDEE---KPVGIG---HSHSNIEYFPKEBVCAGKKGKQKVTGEBELTNVALLKKNSTEN 610
QY      700  NKNF 703
        ||:||
Db      611  NQNF 614

```

RESULT 9

hypothetical 92.4K protein - Streptococcus agalactiae
CISpecies: Streptococcus agalactiae
CDate: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 21-Jul-2000
CAccession: T46758
RSPilebeerg, B.; Rodczinski, E.; Martin, S.; Weber-Heyemann, J.; Schnitzler, N.; Luett
infect. Immun. 67, 871-878, 1999
A>Title: Lmb, a protein with similarities to the Lrai adhesin family, mediates attachment
A|Reference number: Z24091; MUID:99115568; PMID:9916102
A|Accession: T46758
A|Status: preliminary; translated from GB/EMBL/DBJ
A|Molecule type: DNA
A|Residues: 1-822 <SPS>
A|Cross-references: EMBL:AF062533; NID:g4249622; PIDD:AD13797.1; PID:g4249624
A|Experimental source: strain R268
CISuperfamily: Streptococcus agalactiae hypothetical 92.4K protein

Query Match	22.4%;	Score 934;	DB 2;	Length 822;
Best Local Similarity	29.4%;	Pred. No. 2.7e-43;		
Matches 271;	Conservative 99;	Mismatches 235;	Indels 318;	Gaps 32.

```

0Y      1 STELGUOARIVKENNRVS YID---GKQATOKTENLTPEDEVSKREGINAEOVIKIITDOG    57
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     22 SYOLGKHMHGLATTKNOIAIYIDSISKGIVAKPRT-NNTMIOISAEBEGISAQOIIVKITDOG    80
           ::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
0Y      58 VYTSHGDHYHNYNGKVUPDAIISEELMKDPNPKLKDIEDIVNEVSGGVYIKVDCKUYVLV   117
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     61 VYTSIGDHIHFNGKVPDYAIIISELLMTDBPNHFGOSDVINIELDSGYVKANONYYVLV   140
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
0Y      118 KDAHAHADNVTRKEEINROKOEHSOH-REGS-----TPRNDGAVALARSOGRYTTDDGY   169
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     141 KPGRKRKNIRTKQOIAEQVAKGTKEAKEKLQAUVALLSKEEVAAVNBEAKQGFRYTTDDGY   200
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
0Y      170 I FMASDIIEPDGDAYIVPHGDHYHYIPXNELASELSLELAEEAFIS---GRGNLSRSRRYLR   226
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     201 IFSPDLTIIDLDADALVPRGHNHNYIPKKOLDSELAAADAVYSOKGRC--ARPSTDYP   258
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
0Y      227 QNSDNTSRTNWVPSVS-NPGTNTNTSN-----NSNTNSQASQSND-----IDSLLK   272
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     259 TPAPRRRKAP-I PVTNPNGGGHQBDNGGVHPARRPRINDASQNKQHDEFKGTFFKELL D   317
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
0Y      273 OLYUKPLSORHVESGVLDPRAOTISRARGVAVPHGDHNHFI PYSQMSLEEERLARIP   332
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     318 OLHRIDLKYRHNEBEGGLIFEPLOVIXSMAGFYVVPRGDHNHII PRSOLSPLEMLAD---   374
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
0Y      333 L RYRSNMHWPMDSRPEDOSPQTREPSPRGOPAPBNLIKIDSSSGLVRKKUGEEGVPEEK   392
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     375 -RYLAG-----QTDNDSD-----GSDH-----   390
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
0Y      393 GISRYVFAKLPSETFVKNLESKUSKOESVSHITTAKEENVAPRDDEIFYDAYNNILTEAHK   452
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     391 -----SKPSDKC-VTHFLGHR-----IKAY-----   410
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
0Y      453 ALFXHKGBNSDFQALDKLLERLNDESTNKECLVDLLAFLAPIPTHPERLKCPNQSIGEYTE   512
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     411 -----GKGLD-----GKP-----   418
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
0Y      513 DEVARIQALADKYTTYSDGYIFDEHDIIISDEGDAYVTPRMGSHWIGKDSLDEKEVAAAQAY   572
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
Db     419 ----YDTSDAYVFESKESIHSVXSGLTAGHGDFHYIGFELBOYELEDVANW       467
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
0Y      573 TKEGC----ILPPSPDADVKANPTGDSAAL-----IYR-----   602
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
468 VKAQOADELVALLDOEGKEKERPLFDJTKVSVSRKYDGKVGVIIMPKDQKDYFARYOLOL, 527

```


QY 555 WIGKDSLDEKVAQAAYTKKGLIPSPSDADYKANPTGDSAAAIYNRKGEKRIPLVRL 614
 DB 1484 IENAGSIDVNTAKONAL---QALIDPIQASTVKTARAEALTEMONKI-----TGI 1532
 QY 615 PYMVEHTVEVKGNLIIP---HKOHYNIKPAWF---DDHYKAPNGYLTLEDLPAITIKYY 668
 DB 1533 INNNETTNEEK-GNDIGPRAAYEEGLNINNAATTTGDTTAKDTAVQKVOQLHA----- 1586
 QY 669 VEHDPERPHSNDGMSASHVIGKKHSDSDPKNPFADDEE-PVEETRP--AEPEVPVETE 725
 DB 1587 --NPVKRP-----AGKKELDQ-----AAADKKTOIEQTPNASSQOEINDAKQE 1626
 QY 726 KVEAOLKEA-----EVLAKYTDSSLKANATETLAGLRNNITLQIMD----- 767
 DB 1627 -VDTELNGAKTNVDQSSSTEYVDNVAKEGAKINAVKTFSEYKQDLAKIEDAYNAKAVE 1685
 QY 768 -----NNSIMAEAEKLIALLKGSNPSVSK 792
 DB 1686 ADNSNASTSSEIAEAKQKLAELKQTADQVNO 1717

RESULT 14

S56271
 Hypothetical protein YFR016c - Yeast (Saccharomyces cerevisiae)
 C:Species: Saccharomyces cerevisiae
 C:Date: 02-Sep-1995 #sequence_revision 19-Oct-1995 #text_change 19-Apr-2002
 C:Accession: S56271
 R.Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Saseana, S.I.; Sasano
 submitted to the EMBL Data Library, May 1995
 A:Description: Analysis of the nucleotide sequence of chromosome VI from Saccharomyces ce
 A:Reference number: S56186
 A:Accession: S56271
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1233 <MOR>
 A:Cross-references: EMBL:D50617; NID:g836685; PIDN:BA09255.1; PID:dl009896; PID:g836771
 C:Genetics:
 A:Cross-references: SGD:S0001912
 A:Map position: 6R

Query Match 4.1%; Score 171; DB 2; Length 1233;
 Best Local Similarity 19.8%; Pred. No. 0.21;
 Matches 189; Conservative 142; Mismatches 402; Indels 222; Gaps 42;
 QY 12 VENNRRVSTIDGKATOKTENITPPEVSKREGINAEQIYI-----KITD-QGYVTS 61
 DB 151 IKETSTNNVAEG-----TENVPP--IKESTGIEVNSPIITRRKKKKKKTNRGRNS 202
 QY 62 HGDHYHYNGKVPYALII--SEELMKDKNYKDKEDI-----VNEVGGYIKVD 110
 DB 203 NADATDTSKQSTLBSLVIGIEVLOEDGS--KNEDIVNTVODEPVAVERKDIRTRNE 259
 QY 111 GKYYVYLKDAHAADNV-----RTKEINRQKOEHSQHREGG--TPRNDGAVALLASQ--- 160
 DB 260 SSDKTFDIDVPRKNDVDETSSKENNINEKKAHTLPRENNELIANBENNAASFQQLER 319
 QY 161 -GRYTTDGYINNASDIIDTGDAYIVPRGDHYHYIPKNELSASELAALAEAFISGRNIS 219
 DB 320 HOLEAGDE---NGQASTKDVSESESLTKNGFNFKENESKHLAKAGEKQTE---SDRDGIS 372
 QY 220 NSRTYRROKSDNTSRNNWPSVSNP-----GTTNTNTNNSNTNNSQAQSDNI----- 267
 DB 373 PSVLAKNQKETETIGEDHDFEOKDEKDEKCKRKLSEVNHENNMSSHNNAGSSDIIPEPTE 432
 QY 268 -----DSLKLQLYKLPLSQRHV-----ESDGLVDPDPAQ---ITSR 299
 DB 433 RETYDETMGPTRKISDNKNLOHGNDISVEVEKEEBEBEBEBEBESTSKYKKEAVTGE 492
 QY 300 -----TARGVAVPHGDHYHPIPYQSOMSELEBRIATIPLYRSN----- 338
 DB 493 QEAVNNNEVSGTEESTSGKEEIMGCD-----EKOSEAGEK--SSIIEISGSANSAKIS 544

QY 339 --HWVDSRPEQPSDQTPPEPSPGQAPAPNLKIDSNSSLVSQLVKRGVGEYFEKGISR 396
 DB 545 KQNLVLEBAEAPLQENKPTVEVGEIDIPDARND-----VEIVAEVKNIIIPEDLEAVK 599
 QY 397 ----YVFAKDLPSETVKNLESKLSKQESVSHTLTAKENVAIR-----DOEFYDK 442
 DB 600 EDQEGEOVLADPEVPAMKDKIYMRGAESISBDMKKQKQGTAEISNEKAKEVEDFARES 659
 QY 443 AYNLLTEAHKALFNK-----GRNSDPQALDKLLERLNDSTTKKEKIVDLLAFLAI 495
 DB 660 AGEVEVSKTPESPFRVVKRTSGRPEDQINERPEVLEKEDVRVPEDEVKPEIATTTEN 719
 QY 496 THPERLGRKNSQIEYTEDEVRIAQLADKYTSDGYIFDEHD-----IISDEGDAYVTPHM 550
 DB 720 SEEDPKSGRVQISTEQAEITQKMDGDSITS---FKBEKPKRFEITQOEDKITGKT 776
 QY 551 GSHWIGKDSLDEKVAQAAYTKKGLIPSPS-----PDADYKANPTGDSAAAIYNR 602
 DB 777 NHEHGEATEAASENSK-ASDVGTAEKIETPSSSVYKQTEEDAEVE-----NSEKTEFIK 830
 QY 603 VKGEKRIPLVRLPYMVEHTVEVKGNLIIP-----PHKOHYNIKPAWPDHTY 650
 DB 831 VRAE--LENLDAPKEAEVTAELINKENEDVEVDTEDEAEVENSEKTEFIK--AELGNL 885
 QY 651 KAPN--GYTLR-----DLPATIKYYVE-----HPDERPHSNDGMSASHVLGKQD-- 694
 DB 886 DAPKAEVTAELINKENEDVEVAATSKEDLETCSPEAFPIPD---GTOTEAVSGKDAE 942
 QY 695 --HSDPKNKFADEPVEETPAPEVPQVE-----TEKVE-AQKAEAVLLAKYTDSSL 746
 DB 943 AVTKEDENMENSKIYALKVDVGDQEDIDINISDEFQVLEPELEKQIKONKGDKEL 1002
 QY 747 KANATETLAGL-----RNNLTQIMDNNSIMAEKLIALLKGSNPSVSKKIN 796
 DB 1003 EVEETEKETSLPDLVBEENIT--BEKKEIKQEBE-VSQLDFNETESISKEAPN 1053

RESULT 15

D84900
 Hypothetical protein At2g46240 (imported) - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: D84900
 R.Jin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.B.; Umayam, L.; Tallon, L.
 eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487; PMID:10617197
 A:Accession: D84900
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1043 <STO>
 A:Cross-references: GB:AE002093; NID:g3702325; PIDN:AAC62882.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g46240
 A:Map position: 2

Query Match 4.0%; Score 168.5; DB 2; Length 1043;
 Best Local Similarity 18.5%; Pred. No. 0.22;
 Matches 169; Conservative 108; Mismatches 327; Indels 311; Gaps 38;
 QY 40 KREGINAEQIYKITDQGYTSHGHYHNGKV---PYDAIISSEELMKDKNYKDKED 96
 DB 271 KRDDVEASB-----SSNEDRKQWQNKQKTYEVPDISMIKSL----- 306
 QY 97 IYNEVKGGVIVKVDGKYVYVLKDAHADNVRTKEE-----NRQKOEHSQ 142
 DB 307 ----TQG-----QDVKEAQONQKKKEBPQVPIFMIPTSYGKKKDVASBS 348
 QY 143 REGGTP-----RNDGAVALLARSQGYTTDDGYIFNASSDIIDTGDAYIVPHD 190
 DB 349 KESSNEGRNLESCPSDLHRNEGQITQAKK-----EGNEPCNVLDABEKSSVINIPVAN 403

```

OY 191 HY-----HYIPKNEISASELAARAEFLSGR-----215
DB 404 HLOEPRNIPVKLSEHNLKRPTEPTKRIAKNEPVKSTKEQSSSSSEASKLPVCLRVDP 463
OY 216 -----GNLSNRYTRRQNSDNT-----SRTNWPSVSNPGTTNTNTS--NN 254
DB 464 PKERNQSKSVSHPRKMEKSKETKIAAPLSKKAESRT--VPEACNVACEDANAEKMAE 521
OY 255 SNTNSQAOSNDIDLLKQLYKLPLSQRHVESDGLVPFPAQITSTAGVAVPHDHF 314
DB 522 GSLNALRTEKGSVES-----NSNLQESNGEIIKPCAKENREOP-----561
OY 315 IYPSOMSELEERIIPLRYRS---NHWPDSREOPSPOPTPEPSPG-----360
DB 562 ---AKSFTBEEAAILIOWMYRGVYRRREPIKIKLKEIA---TVAEQGDVKKRIEAL 615
OY 361 ---PQAPNULKIDNSLSVQVLRKVGEGVPEEKISRYVPAKDLPSFTVKNLESKLS--416
DB 616 STDQHIIEKEKIVVNGELVWNLILKLD---AVEGLHPSIREFKALATE-LSSIQDKLDSL 671
OY 417 KOESVSHLTAKKEVAVAPRDOEFVDAKAYNL---LTFAHKALFNKKGNSDFOALDKLE 472
DB 672 KNSCASAKEAKVEQVEIKSQP-SDSPVNLHESQLTEENKAV-----712
OY 473 RLNDSTNKEKLVDDLAFAPITHPERLGKPNQIEYTEDEVRIAQDLADKYTSDGYIF 532
DB 713 ---SDTLLEKTLR-----LSPEEH-----PMSVLNRTDER---QAESAAETEGY--751
OY 533 DEHDIISDEGDAYVTPHNGSHWICKDSLQKEXVAAQAYT-----KEKGILPSPDA 585
DB 752 -----GLFETLATDSQATENAAAASSTTIPKIGEVETVPGND--791
OY 586 DYKANPTGDSAAIYNRKXGKRI-----PLVLLPVTVEHT-----VEV---624
DB 792 ---PSADNGMTVTNVEENKAMVAVESLEEPINELPQVEETETNSLRDPENASEVSEA 846
OY 625 -----KNGNLIIPHKD---HYHNIKFAMFDDHT---YKAPNGYLTEDLPA-----663
DB 847 ETNSEENENRKGEDIVLHSEKNVLSLSPGVVIDEETOPLSQDSSSYTRGKNTAMD 906
OY 664 -TIKYVHPDERPHSNDGWNASEHVLGKKDSEDPMKFADEEVE-----ETPAE 716
DB 907 KTAQSEETEVDSHPNNSKIGIQOTSEPODEKQOS--PETEVIVKQPLTEVILNEQAP 964
OY 717 PEVPEVETBEVBAQCKEAEVLLAKYTDSLSKANAFTLAGLRNNTLQIMDNNGSMAAE 776
DB 965 PEITEPGISKETKLMENQRFKEMTETLVKAGREQ--LEVISKLTSSVKLSLEKLSHK 1022
OY 777 KLALLKGSNPSSVS 791
DB 1023 KTOIRRRASKPMWSVS 1037

```

```

RESULT 16
G89957
hypotheical protein [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: G89957
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc
ma, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattoji, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1235-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: G89957
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-891 <KUR>
A:Cross-references: GB:BA000018; PID:g13701526; PIDN:BAB42820.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:

```

```

A:Gene: SA1552
Query Match 4.0%; Score 165; DB 2; Length 891;
Best Local Similarity 19.0%; Pred. No. 0.27; Mismatches 323; Indels 292; Gaps 48;
Matches 177; Conservative 140; Mismatches 323; Indels 292; Gaps 48;

```

```

OY 7 YQARTVKENRVSYIDKQATQKTENLTPEVSKREGIMAEQIVIKITDQGYVTSHGDIY 66
DB 36 HQQAQENVTNSDKISENNQNNATTTQOPKQNTQO--PATQPV--ITAKVYPA-----86
OY 67 HYVNGKVPYDAIIEBELMDPNYKLDKEDIVNEVKGIVY---KIDGKYVY--LMDA 121
DB 87 -----DESLKDAI--KDPALENKEHDIGPREQVNFOLDKNNETQYVHFPSIDPA 135
OY 122 HADNRTKEI-----NRQK-----QESQREGCTPRNDG 152
DB 136 DVTYTKKAEVELDINTASTKKEFEVENQKLPVRLVSVSPVEDIAYIR---FPSD 192
OY 153 AVALARSQGRYTTDDGY-----IFNASDIE-DTGDAYIIVPHGHVHYIIPK 197
DB 193 TQEL-KIVYSTQIDQGETNYDYTKLVFAKPIYNDPSLVKSDTNDAYVT-----240
OY 198 NELSASELAARAEFLSGRGLSNRTYRRQNSDNTSRTNWP---SVSNPG-----245
DB 241 NDQSSDAS-----NOTNTNT--SNQNTSTNNANNOPOATTNMSQAPKSSANA 289
OY 246 -----TNTNTSNNSNTNSQASQNSDIDSLKQLYKLPLSQRHVESDGLVFD-----292
DB 290 DQASSQPAHETNSGNTNDKTNSSNSD-----VNOQTPPADESIQDAIKN 336
OY 293 PAQITSTARGAVAPHGDHYHFIPIYSQMSLEERIIPLRYRSNHWPDSREOPSQ 352
DB 337 PAIDKEHT-----ADNWRPIDQMKNDGER-----QFYHA-----369
OY 353 PTPEPS-----PPQAPNULK-----IDSNSLSQVLR-KVGEVYVEEKIS 395
DB 370 STEPARATVIFTKTPVLELKTASTWKKEFEVGDGKLPELVSYSDSKQVAYAIRPVS 429
OY 396 RYVPAKDLPESTVKNLSKLSKQSVSHLTAKKENAVAPRDOEFPYD-KAVN--LTFAH 451
DB 430 N--GTRDV--KNVSSIEGENIHEDYDTLMAVFOPIITNPPDDYVDEETVYLQKLLAPH 485
OY 452 KALFXNGRNSDFQALDKLERLNDE--STNKEKLVDDLAFLA-----PIHPERLGKPN 505
DB 486 KA-----KTLERQVVELEKQELPKYKAEYKKLQDQRYVELADQVSAVTEFENVPTN 541
OY 506 SQIYTEDEVRIAQDLADKYTTS--DGYIFDEHDIISD--BGDAYVT--PHMGSHWIGKDS 560
DB 542 DQLTDLQEAHFVVESESENSESVMDGFV--EHPFYATLNGQKQVVMKTKDSDYV--KDL 597
OY 561 LSDKEVAAQAYTKEG---ILPSPPADVKAAPTGSAAIYNRKXGKRIPLVRLPY 616
DB 598 IVEGRVYTVYSKDPKNSRTLIIFYIPD-----KAVYNAL--YKVYVANIGY 642
OY 617 MVEHTVEKGNLIIPKDHNYNI-----KFAWFD--DHTYKAPNGYLTEDL 661
DB 643 EGQHVHRIINGDINTXDDOTSQNNITSEPLNVQTOGEGKQVADTVDAENSSSTATPKDASDK 702
OY 662 FATIKYVHPDERPHSN--DGWGNASEHVLGKKDSEDPMKFAKDEEVEETPAEPE 718
DB 703 ADVIE-----PESDVVDAQANNIDKQVHDVHIDMSDNNHFD-----741
OY 719 VPQVETKEVQAEVLAKVYDSSILKANATELGLRNNL-----TL 763
DB 742 -----KDLKEMPTQIAKQDTRNV--DNSVGSNSVDTDXSNKKQKVIOLA 787
OY 764 QIMDNNSIMAEAEKLLALLKG-SNPSSVSKEK 794
DB 788 HIAQKNHTGKAALDLVQKQYNNVTADVTDK 819

```

```

RESULT 17
A44361

```

amiloride-sensitive sodium channel α px protein - African clawed frog

M/Alternative names: apical plasma membrane protein

C/Species: *Xenopus laevis* (African clawed frog)

C/Date: 30-Apr-1993 #sequence_revision 19-Oct-1995 #text_change 22-Jun-1999

C/Accession: A44361, S25517

R/Status: O.; Verrey, F.; Keyman, T.R.; Benos, D.J.; Rossier, B.C.; Kraehenbuhl, J.P.

J. Cell Biol. 119, 1497-1506, 1992

A/Title: Primary structure of an apical protein from *Xenopus laevis* that participates in

A/Reference number: A44361, MUID:93107151, PMID:1334959

A/Accession: A44361

A/Molecule type: mRNA

A/Residues: 1-1420 <STA>

A/Cross-references: EMBL:Z14997, NID:G64551, PID:CAA78718.1, PID:G64552

A/Experimental source: A6 cells

A/Note: Sequence extracted from NCBI backbone (NCBI:P121141)

C/Complex: This protein is part of a large molecular complex.

C/Function:

A/Description: may be the amiloride-sensitive component of the amiloride-sensitive sodium

C/Superfamily: amiloride-sensitive sodium channel α px protein

C/Keywords: glycoprotein; membrane-associated protein; sodium transport

F:119,462,481,503,660,664,988,1038,1211,1273/Binding site: carbohydrate (Asn) (covalent)

Query Match 3.8%; Score 161, DB 1; Length 1420;

Best Local Similarity 19.8%; Pred. No. 0.89;

Matches 180; Conservative 111; Mismatches 318; Indels 302; Gaps 44;

42 EGINAEQIVKIDTDOGVTSQHDY-----HYNGKVY-----DALISE 82

26 ERISPRSMITLVDSAYSSFGSSVPEYONSFOHDCCHNDQLYMSEYRAIYNPS 85

83 LLMKDNVYKIKDEIDVNEVKGVIKVDG-----YVYLDAADADNRKTR-- 129

86 LLDKDGIV-----NDIVSE-HGSSKVALSGRSSSLCSDMTTSVHRTSPAKLDVYNNLDS 140

130 -----EELN-RQKQHSQHSRREGTFRND--CAVALARSQGR-----YTTDDGTFNNA 173

141 EKNIYDPIPMKQKORPNHKAAGLORNSPTGINSLOEKENQLYNSPMELIKDNYFGRS 200

174 SDIIEDTGAIVYIPHGDIH-----YIPKELASSELAAEAPLISGRGNLSNR--TYR 225

201 LDVLADGD--IMTDDSTQNALYFPQNDPOYR--NTQYFGANRMSKEQFKYNDVQ 253

226 RONSNDTSRTNWVSVSNPGTTNTNTSNNTNSQASQSDNDLSLKQLYKPLSQRYHVE 285

254 KSNBENTERDG--PYLTQDQGF-----VQGVASDVATRSKNI-----RRLSK 294

286 SGGLVFDPAQITSRTARG--VAVPHGDHYF-----IPYSQNSELEERTARI-- 331

295 KSA---SGKIVAHDSQSCWIMKPKDTPSFNSEGTITDMQDNRQMDIRKSRISTRA 350

332 --PLRYRSHWVVDSPDQSPQPTPEPSGQPAENLKDIDSNSLVQLVKRVGSGYVF 389

351 SGLSYVESMEDV-----SGEPLKAMSKNVDTLS----- 381

390 EEKGISRYVFAKDLPESETVKNLESKLSKQSVSHITLAKENVAAPDOEFYDKA--Y 444

382 -----FOKDATVKSIPILSLOLOQEKCKSHPL-----SDLNCEKITASTPMY 425

445 NLTLEAHKALF-----XNKGNSDFQALDKLERLNDSTENKEKLVADLLAFLAPITHE 499

426 HLAGGSHSAFIAVHNHTNPAQOEKLEKSTLERMNNIS-----VLQJSEPRPDH 476

500 RLGKPSQLEYTEDEVRIOADKYTTSQGYIFDEHDIISDEGDAVTPHMGSHHIGD 559

477 KLPPKNS-----LTQLADLHDSVSG-----NNSGNLNSABE 508

560 SLS-----DREKVAQAQYTKEKGI-----LPPSPADADYKANP----- 591

509 SLMDNDIETKLKVAQKVLARETSFKRDLQMSLPCRKLNDPKKPTIDHRSYSSSANE 568

592.-----TGDSAAAIY-----NRVKGKRIPLVRLPYMVEHTYEVKGNLIIPK-D 635

569 SAYLQTKNSADSSYKKDDTEKAVATRIIGRKRI-----TKEQKKLCYSEPEKLD 617

QY 636 H--YHNIRKAPMDHTYKAPNGYTTLEDLPATIKYVVEHPDEPRHNSDGNASEHVLGKK 693

Db 618 HIGIQKSNAMKEEFTFANRRMSDSDISANIKYLE-SKETNNS--SNISKTELKQI 673

QY 694 DH-----SEDPKNKFADEEVEETPA-----EPEVPOVETKEVAOLK---- 732

Db 674 QHNALVQWERKTQNRPNNSN---POVQMERSTSLGPNVNEWISSETSSDASQKYLRR 730

QY 733 ----BAEVLIAVY-----DSSLKANATETLAGIRNNLTQINDNNSIMAEELAL 781

Db 731 RSAGASSSYDATVATVWNRFRKTSPLGRSAEXTAGVQK--TFSDQRTLDGSGQEH-- 783

QY 782 LKGNPSSSVSK 792

Db 784 LGGSSPSSLQK 794

RESULT 18

T22298

hypothetical protein F46C3.3 - *Caenorhabditis elegans*

C/Species: *Caenorhabditis elegans*

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C/Accession: T22298

R/Cottage: A.

submitted to the EMBL Data Library, November 1995

A/Reference number: Z19543

A/Accession: T22298

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-2810 <WIL>

A/Cross-references: EMBL:Z66563; PIDN:CAA91469.1; GSPDB:GN00028; CESP:F46C3.3

A/Experimental source: clone F46C3

C/Genetics:

A/Map position: X

A/Intons: 67/2; 172/3; 204/2; 310/3; 366/2; 431/2; 684/3; 739/3; 786/3; 823/2; 880/2; 1958/3; 1999/1; 2078/1; 2117/3; 2159/2; 2220/1; 2269/3; 2306/1; 2399/2; 2444/1; 2488/3;

Query Match 3.8%; Score 159, DB 2; Length 2810;

Best Local Similarity 18.5%; Pred. No. 3.1;

Matches 183; Conservative 137; Mismatches 346; Indels 322; Gaps 46;

19 SYIDGKQATQKENTLPDVSREKGINAE--QIVIKITDOGVTSQHDHYHY-NGKVP- 74

Db 1542 STV--PQYQIQYCPPEPILSPQSVRSVPPMAMPVHMDGSTRSDYRIKMGSEVPS 1599

QY 75 -YDAI-----ISEELMKDPN--YKLKDEIDVNEVKGVIKVDGKY 113

Db 1600 QYSTIRNMVPRHGDVDFLDVAVFDQVLSKDEQGAHNSNQNLNTIGG- KLRPEGY 1658

QY 114 -----YYLTK--DAHADNVYKEEINRQOEI-----SOHREG 145

Db 1659 EPPVQTVSEVPPRYPTLRRVDSPLRSRAKSLPRIISPRHEHFRVRPHSRNSYSNESTSS 1718

QY 146 GTPRNDGVALARSQGRYTTDDGYIFNADIEDGDAIYV-----HGDHVIYIPKEL 200

Db 1719 DDQMYRTSRRSRSIPRFHNSNGYVDPQPV-----YMMVQVNGHEMILLSP-- 1768

QY 201 SASLELAAPFLSGGNLSNSTRYRONSNDTSRTNWVSVSNPGTTNTNTSNNSVTNSQ 260

Db 1769 VASEHKAO-----SRHRTDR-----SHKHNV-----SGVEYMAK 1800

QY 261 ASQSNDI-----DSSLKQLYKLP-LSGRVHESDGLVDPQAQITSRTAGVAVPHG 309

Db 1801 RSPSVDLKLPMSRRTPDAMVQYVVRPLAOSP-----GROTSEFEFSALPRG 1851

QY 310 DHYHPIPSQMSLEER-----IARLIP--LAVRSNHV--PDSRPEQSPQPT----- 354

Db 1852 D-----SRAREKENPLARQIPRAYVQNGVAVPPSPSTRAPATSGDRRG 1901

QY 355 ----- 354

```

Db      1902  LNRLEQESYTEBAVNTKNNSEYLSPHRENLDQKAVIRHEKETAKNAIMLSDRLSPV 1961
QY      355  PPSGPGOPAPMLKIDSNSLSVQLVRKYGEGVEE-----KGISRYVFAKD----- 402
DB      1962  PPTPPPPPP-----VREKWIIDSVERDPVYQGRIRGSRKREPLVP 2004
QY      403  LPSEIVKNLESKLSKQESVSHLTAKKENVAP-----RDOFYDKAYNLTLEAHKLFXXK 458
DB      2005  QPPRPVNAEKPAVNFVKAFWKLTIRKEMFYPGEVANDIQIIDQVPAQVLECKKSYPYR 2064
QY      459  GRNSPFOALDKLLERLN---DESTNKEKLVDDI-----LAPLAPL---TPERLQKPN 505
DB      2065  IREQRKQVETLRQVEIIPPSDIINNQSNIHPDKAVAVIELARLWPLYFNQVVEVEKSPD 2124
QY      506  SQIE-----YTEDEVRIA-----QLADKYTSDGYTFDEHDIISDGDAAVTPHMGSHHI 556
DB      2125  ESVSTIFALSEHGIRLIYHTPHDLENPKIQDPPRETTADVLSLEANDILSVHVN----- 2180
QY      557  GKDSLSDKEKVAQAAYTKEK-----GILPSPDADVKAN-----PTGDSMA 597
DB      2181  -----EDENAYSAVRIKTNQAPQIKTLDRCLSGSVVPRKRFVLARTGRETTPVGN--- 2232
QY      598  AIYVNVKGEKRIPLVRLPYMVEHTVEVKNGLIIP---HKDHYNIKFAMFDHTTYKA 652
DB      2233  WLYGKIEN-----RFGLAQYVDSTGDVNPPIRHETSEDREVRVF--FDEDEVPS 2283
QY      653  PNYGTLEDFAFIKYVVEHDERPHSNDQGNASEHVLGKHSDRP--NKNFKADEPVE 711
DB      2284  SERVTMID-FAT-KTFRKPKDKKQETWAMEDISQV---RFSEKPIQSOLADLQNEE 2337
QY      712  ETPAPEVPOVETEKEVAQLKEAE-----VLLAKVYDSSIKANATETTLAQRNNLT 763
DB      2338  SKYAVETFAIMKFMGDEPLKKSSEWTVDFVFLVILICHQPTLR---DEVYQQLKQTS 2394
QY      764  QIMDNNSIMAEKLLALLKGSNPSVS 791
DB      2395  NISQKPNASALRAMRLTLITTAFFPSSLT 2422

```

RESULT 19

```

T18444
Hypothetical protein C0385C - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
C:Accession: T18444
R:Lewson, D.; Bowman, S.; Barrett, B.
submitted to the EMBL Data Library, August 1997
A:Accession: T18444
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1650 <LAW>
A:Cross-references: EMBL:298547; NID:e1325376; PID:e1427940; PIDN:CAB11112.2
A:Genetics:
A:Map position: 3
A:Introns: 1597/3; 1625/3
A>Note: C0385C

```

```

Query Match      3.8%; Score 158; DB 2; Length 1650;
Best Local Similarity 17.7%; Pred. No. 1.6;
Matches 146; Conservative 127; Mismatches 311; Indels 242; Gaps 34;

QY      64  DHYHYNGVPPDAIISEELMKDKNYKAKDDIYNEVGVGVIVDGKYYVYLADAHHA 123
DB      11  DNVHYTKKNV-----IRNYDDHNITKANSHLKQSKY-----TTHN 47
QY      124  DNVV---TKEINROKQESHQREGTPRNDGAVA-----LARSQGRYT 164
DB      48  DEKQWITSNNNNNNIQTAA-----NVKDDKSVENKTTTLINKDINETINSKQDKYIT 102
QY      165  TDDGTFVNASDIIEDTGDAIVPHGDHYIYIKNELISASELAABAFLSGRGNLSNRTY 224
DB      103  KYNCLNFNKSVDVMD-----YVRYSKKEDSSNNNGNNGNNNNNGNNNNNGNNNM 150

```

```

QY      225  RRONSDNTSRITNWPVSVSPGTTNTNTSNGNTNSOASQSDNIDSLKQLYKLPLSQRHV 284
DB      151  NGNNNNM-----GNMNNNGNNNNNGNNNNNDNNNNNG-----NNI 185
QY      285  ESDGLVFPDAQITSRIT-ARGVAVPHGDHYHIFPYQSOMSELEBRIATIIPLRYSNHWV 342
DB      186  MNGNNIMNGNNNNNNNNNNNNNNNNNNNNNN-----NINNNNNNR-----NNITNNHIIH 234
QY      343  DSRPQSPQPTPSPSPQAPNL-----KIDSNSLSVQ----- 378
DB      235  HISNTSLIENKKKEENIFPSANLYKKNVKGSSDTFDLLYRKRIKDKDNLKSKKRP 294
QY      379  -LVAKVEGYVFEKGISR-----YVPAKDLPESEIVKNLESKLSKQ 418
DB      295  NYLFRKQGEHFVNKENVQNNNIIDDDDDDDNDHNDVNVVYDYKENEMENKKNKKSVE 354
QY      419  ESVSHLT--AKKENVAPRD--OEYDKAYNLTLEAHKALFXNGRNSDFOALDKLLER 473
DB      355  DGLHVLVELRNKNLNVNDNIINNSFEKNILYIKTSDSLNENYERKIKYEINK----- 410
QY      474  LNDESTNKEKLV-----DDLAPLAPITHPERLQKPNQIETDEVRIAQLADKYTSD 528
DB      411  --EESYNNREYVHFRKND--SSIKKNNSSCLDEQKKTY--RYTII 454
QY      529  GYTFDEHDIISDEGAYVTPHMGSHWIGKDSLDEK-----VAAQATKEKGIPLP 581
DB      455  QKRYNFD--RDNNNAYI-----KDTTKKEKGYLYNMIYQSEEL--KTYGSNNK 500
QY      582  SPDADVKANPTGD-----SAAAIYVNVKGEKRIPLVRLPYMVEHTVEVKNG- 627
DB      501  MDEMEIYQHTNDPFINENLNKKIYFDDEGYDEKKKKKLD---DHIIYQOKEKNNI 556
QY      628  NLIIPIKHQHYNINIFAMPDHTYKAPNGYTLDELFAFIKYVVEHDERPHSNDQGNASE 687
DB      557  NDIL--KDHILN-----DKETKEKKEIEIEE--EKXKIEIEEKKKIE---IEE 601
QY      688  HVLGKHSDEDPNNKFKADEEPVEETPAPEVPQVETEK-----VEAQLK 732
DB      602  EKKKIEIEEKKKKKIEIEEKKKKIEMEERKNKIDDKKNTYANDKIIISHIDVNNCAIK 661
QY      733  EAEVLLAKVYDSSIKANATETL-AGLRNNLTQIMDNNSIMAEK 777
DB      662  -IDALLDHIIEKKTKYGHKEINLYKEIKNEYQKMLDENSEIMLEHEK 706

```

RESULT 20

```

S73361
dnaJ homolog protein C09_orf910 - Mycoplasma pneumoniae (strain ATCC 29342)
C:Species: Mycoplasma pneumoniae
A:Variety: ATCC 29342
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 10-Dec-1999
C:Accession: S73361
R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirk, E.; Li, B.C.; Herrmann, R.
Nucleic Acids Res. 24, 4420-4449, 1996
A>Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumoniae
A:Reference number: S73327; MUID:97105885; PMID:8948633
A:Accession: S73361
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-910 <HIM>
A:Cross-references: EMBL:AE000004; GB:U00089; NID:g1673671; PIDN:AA895683.1; PID:g167368
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
A:Genetics:
A:Superfamily: Mycoplasma heat shock protein dnaJ homolog C09_orf910; dnaJ amino-terminal
F:7-71/Domain: dnaJ amino-terminal homology <DN>

```

```

Query Match      3.8%; Score 157; DB 1; Length 910;
Best Local Similarity 19.0%; Pred. No. 0.77; Indels 254; Gaps 44;
Matches 161; Conservative 127; Mismatches 305; Indels 254; Gaps 44;

QY      66  YHYNGKVPDAIISEEL-----LMKDP-----NYLKXDEDIVNEVKGGVYIKVD--GKYY 114

```

Db 34 YHEDRNKADPAQAQIFAEINEANDVLNPKKRAYDYKXGHDGVN--EGGFAPQADVPDSFF 92
 115 VYLKDAHAADNVYTKKEINRQKQESOHREGGTPRNDGAVALARSGRYTDDGYIFNMS 174
 Db 93 EREKSGADNL---SESNTKKKEKTKTKKKW-----FWKSKQESTSTTYY--A 140
 175 DIIEDTGADYIVPHGDHYHYPKNELSASELAABAFLSGKGNLSNSTRYRONSNTSR 214
 Db 141 DV--DAGLEDYPRQSGYPPDIIPVDARIEVDQS--AYADDIPVDAGMWMQNAEYANSA 197
 235 TMMVPSVSPGTTNTNTSNNSTNSQASQSDIDSL----- 271
 Db 198 SEIIPV-----DAGLADEFNTSSAAPQASDWEAMIGMPEYGFPAAGEMNKKGFDEA 251
 272 -KQVLYPLSQRHVESDGLVFDPQQTSTRARCVANPHGDHYHFIYSQSELEERIARI 330
 Db 252 GQVWVLEETEPSSVSNDETITTSDAVTAAT-----VEETDQ----- 288
 331 IPLRYSNHVPDPSPR-----QSPQPTP-----EPPGPR-----OPAPNLKI 369
 Db 289 -----DSMTANSAPEDVDVETPVELQPTETEPITLSSFEVEAPASVIEPTPEIE 341
 370 DSNSSLVSQLVKRVGEGYVFEKGISRYVPFADLPSETVKNLESKLSQESVSHLTAKK 429
 Db 342 TTSAVEMDASVK---ADVSDADATNEPTEDTISEPQ--ETDAAALEINHNT--TADL 394
 430 ENVAPRDPQFYDAYVLLTE--AHKALEKXKNGNSPQALDKLELNDESTKEKLVND 487
 Db 395 E--PAEV---SATNDLEQDVVEKVFSEPESTVTAAADPVEGATESTNGFKF--N 446
 488 LIAFL-----APITPERLGKPNQI-----EYTEDVRIQOLA----- 521
 Db 447 FSSFVLSQNPQCTTHHEEDAAAPRYVDERTSGSTAPETIASVLELTAALINP 506
 522 -----DKYTTSQGYIFDEHDIISDEGDAYVTPHMGSHWIGKDS 562
 Db 507 ATFFVEYLQPTKTTVDKLEPTVAKPTVSDSENSVAPEPEVAGPEQFTS--W--KPAIS 563
 563 DKEV-----AAQATKEGILIPSPDA-----DYKANPTGDSAAA-----IY 600
 Db 564 ETEIELTAVEPASETQTLAEDVTSPTVPTATAIPASINAVPTAPVETFEAAYDFLK 623
 601 NRVKSGKRIPLV--RLPYMEHTVEVKNGLIIPKHQVHNHKKPWFDDHTYKAPNGYTL 659
 Db 624 EAAKIAQLPLVTVVEQIDGT---DPSLL-----TQWDEYIEKTRK--LTH 665
 660 DLEFATIKYYEHDEPHSNDGNGNASEHVLGKHSE--DPKNKFADEPEVEETPAEP 717
 Db 666 KLFLLT-----EQLP-----FIVKTDQFEIYDPR---LDEHVNLLYTE-- 700
 718 EYVQVTEKVEAOLKEAEVLLAVTDSLKANAETELAGLRNNLTQINDNNSIMAEK 777
 Db 701 HYPQIIF--LINEQLKEIR--YTRKLVDPQQTVTTT-----SITLEVQLSHKSQTEA-- 748
 778 LIALLKG 784
 Db 749 -IAIFKG 754

RESULT 21

S55101
 hypothetical protein YMR219w - yeast (Saccharomyces cerevisiae)
 N/Alternate names: hypothetical protein YMR261.13; hypothetical protein YMR959.01
 C/Species: Saccharomyces cerevisiae
 C/Date: 08-Jul-1995 #sequence_rev150n 09-Mar-1996 #text_change 29-Oct-1999
 C/Accession: S55101; S57587
 R/Bedman, K.; Brown, D.; Bowman, S.
 submitted to the EMBL Data Library, June 1995
 A/Reference number: S55089
 A/Accession: S55101
 A/Molecule type: DNA
 A/Residues: 1-711 <DED>

A/Cross-references: EMBL:Z49809; NID:g854459; PIDN:CAA8934.1; PID:g854471; MIPS:YMR219
 A/Experimental source: strain AB972
 R/Skelton, J.; Churcher, C.M.
 submitted to the EMBL Data Library, June 1995
 A/Reference number: S57587
 A/Accession: S57587
 A/Molecule type: DNA
 A/Residues: 608-1658 <SKE>
 A/Cross-references: EMBL:Z49939; NID:g887599; PIDN:CAA90190.1; PID:g887600; MIPS:YMR219
 C/Genetics:
 A/Map position: 13R
 Query Match 3.8%; Score 157; DB 2; Length 1658;
 Best local similarity 19.7%; Pred. No. 1.8;
 Matches 176; Conservative 119; Mismatches 338; Indels 262; Gaps 41;

QY 21 IDGKATOKTENLTPD--EVSKRE---GINAEQIVIKITDGYVTS-----HGDHYH 68
 Db 193 LEGKQSLIKDFLNDVEYELSEEKNSDQSSPSIMI--LSDEYALEGALQVSNDEYAE 251
 69 YNGKPYDAIIEBLMKD-----PNY-----KLKDE--DIVNEVK----- 102
 QY 252 EEGQVERKNIQGEQAVENATQISSDSSEGGQYSEGVEMLEDDIDVESDAEKESQGA 311
 Db 103 GGVIKVDGKYYVYLKDAHAADNVRTKEINRQKQESOHREGGTPRNDGAVALARSGQR 162
 312 EGTSHSVDSKYM---QRTNTYKIPLYEKESDEKHQO-----R 349
 QY 163 YTTDDGYIFNADII-----EDTGAYIVPHGDHYHYPKNELSASELAABAFLSG 214
 Db 350 YEDGAFDQGSVNISVDESEDESOAESYSANAEVYHH--NEHLEDKEL--IEDIES-- 405
 215 RGNLSNSTRYRQNSDNTSRTWVPSVSPGTTNTNTSNNSTNSQA-----S 262
 406 ---DSSESQSAQESQEGSEDFEYKMKNEKTSERTSRTSROGFAKDAYTKNKVQ 461
 QY 263 QSNDDI-----SLKQLY-----KLPLSGRHVESDGLVDPDA-----QITSRT 300
 Db 462 QENDEPEKDDIIRSLDKNPFNGNNKSEYENULENE--TDPAIVEENQINDVEGYD 518
 301 ARGVAPPHGDHYH-----FIPYSQ-----MSELEERIRIIPLYR--RSNHV 341
 Db 519 YNGKSVESDLHHSPLNDLYLARAMLQFOQRNSNCPQKEQVSE---SLGHSNGSN 574
 342 PDSRPEQSPQPTPEPSPEGPAPAKIDSN--SSLVSQLVKRVGEGYVFEKGISRYV 399
 Db 575 LSGRSLDESEEQIPLKDPFTGENNNNLKTRDGLSSVEIEVEKVB----- 620
 400 AKDLPSETVKNL-----ESKLSQESVSHLT-----TAKKE 430
 Db 621 -KKLGSTKEKELVPLSTDTTINNSSLGNEDSIYYSLDDADAISENLTDVPLMEIKTPPY 679
 431 NYAPRDPQFYDAY--NLTLEAKALFXKNGNSPQALDKLELNDESTKEKLVNDL 488
 Db 680 EVVISESVYSSTSYEDNTYAMPQVEYTSPPFNDDPNS-----LNDYEXKHDLKST 732
 489 LAFIAP--ITHPE---RLGKNSQIEYTEDVRI-----AQLDKYTTSQGYIFDEHD 536
 Db 733 LALAPAFPTKDAEFVEAGVTKSLTSGHTNIPHTSKETKQVSDLDSTENVTF--ENE 791
 537 IISDE--GDAYVTPHMGSHWIGKDSLDPKEKVAQAQYKEKGILIPSPDAKYKNPTGD 594
 Db 792 NMGDEKNQSNKPPGVAINTDSTEDNTEKYPFSAINTY-----NVTSD 835
 595 SAAATYNRVKGKRIPLVRLPYMVEHTV-----EVKNGNLIIPKRD--HYANIKF 642
 Db 836 SSGCE-----DIETASVNEENLKYCKDMNEAMSSGDECVKQNDGSKTQISF 884
 643 AMFDHTTYAPNGYTLIEDLFAITIKYYVEHPDRPHSNDGNGNASEHVLGKQD--HSDPN 700
 Db 885 STDSPNFQESNDNT---EFSSTKYKVRNSLDEDESLKKELTKAELVAVDKLDEESEDSEY 941


```

Db      1339 MEDNEKQOPLPETRKEEBEELKEDTSYDGEKMEQYOTEAVELLQEDVDIGSLBESQOQ 1378
      753 TLAG---LRNNLTLOI---MDNNSIMAE-----EKLIALKGSNP-----SSV 790
      1379 NANNETROKSETLOVSKYRIDRMSNLKKAIIILRRVKALENVRKNPREPRFLPNPEV 1438
Qy      791 SKEKIN 796
      1439 EAEKVN 1444

```

RESULT 24

probable bZIP transcription factor [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C:Accession: G84598
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 Neus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MID:20083487; PMID:10617197
 A:Accession: G84598
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-519 <STO>
 A:Cross-references: GB:AE002093; NID:g4582442; PIDN:AAD24827.1; GSPDB:GN00139
 C:Genetics:
 A:Gene: At2g21230
 A:Map position: 2

Query Match
 Best Local Similarity 3.7%; Score 156; DB 2; Length 519;
 Matches 125; Conservative 84; Mismatches 176; Indels 260; Gaps 28;

```

Qy      246 TTNNTSNNSTNSQASQNDIDSLKQLPLSORHYESDGLVDPQAQTSRTARGVA 305
      7 TTDTNMQRVNSSSGSSSS-----IPKHLHL-----NPLALS----- 41
      306 VPHGDHYH-----IYSGM-SELEERIRAI--PIRYSNHWVDSRPE 347
      42 -----HHFRHPFTGAPPPPIPIPSQIPATLQRHSRSMQSPFSFSLPLPNS 96
      348 QSPQPTPP-----SPGQAPAPNLIKINSNSLVQLRYKGEVFEKGSR----- 396
      97 AFSVSVSVEEKTGAGFSPPLPSPTMCHSSSS-----RMAGDGENLPKRKSHRSNSD 150
      397 -----YFAKDLPESETVKNLESKLSKQESVSHLTAKKENVAPRQEFYDKAYNLTBA 450
      151 VTFGFSMMQSQKQSPPLSLERSISGEPTSDMSNLVKE--PR-EGFY----- 196
      451 HKALEFKGRNSDFQALDKLERLNDESTNKEKLVDDLLAFLAPITHPERLQKPSQIEY 510
      197 -----KGR-----KP--EVA 205
      511 TEDEVRIALADKYTTSDGYIPDEHDIISDEGDAYVYTRMGSHWTKGSLSDKEVAAQ 570
      206 AMDDVFTAYMN-----LDNIDVLNSFG-----GEDGKNGNENVEEMESSRGS 247
      571 AVTKEGILPPSPDADVKANPTGDSAAATYNNRYKGEKRIPLVRLPYWVHTVEVKKGNLI 630
      248 GTRKTING--GSSSDSGDSSASAGNVVALSSSSSGYK-----RAGGDI 289
      631 IPHKHVYNI-----KFAWPDHTYKAPNGYTLDELPAIKYVVEHPDERPHS---- 678
      290 APTGRHYRSVMSDCMGKLNFGDESSLKLP-----PSSSAKY 327
      679 ---NCGWQASHEVL--GKKHSHEDPNKFKADEEVEETPAPE----- 718
      328 SPTNSEGNSAVSVFEGNSEFTAAEMKKIADCKLAELVMADPKVKRI LANRVSAAAS 387

```

```

Qy      719 -----VPOVTEKVEAQLKEAVLLAKVT-----DSLK--ANATETLAGL 757
      388 KERKTRVAAELE-HKVOTLQTEATTLTAQLTFLQRDSMGLTRQNSBELKRLQAMEQOQL 446
      759 RNNLLQLMDNNSINIAEAEKLLALL-----KGSPPSSVSKKIN 796
      447 RDALSEKLINE-----EVORLKLVIQEPNRRSGSSSSSESKMSLN 485

```

RESULT 25

hypothetical protein C0105w - malaria parasite (Plasmodium falciparum)
 C:Species: Plasmodium falciparum
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jun-2000
 C:Accession: T18416
 R:Lawson, D.; Bowman, S.; Bartell, B.
 submitted to the EMBL Data Library, July 1997
 A:Reference number: Z18934
 A:Accession: T18416
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1338 <LAW>
 A:Cross-references: EMBL:Z97348; NID:e1323671; PID:e1323672; PIDN:CAB10568.1
 C:Genetics:
 A:Map position: 3
 A:Note: C0105w

Query Match
 Best Local Similarity 3.7%; Score 154.5; DB 2; Length 1338;
 Matches 153; Conservative 122; Mismatches 287; Indels 357; Gaps 35;

```

Qy      11 TVKENNRVSYIDGKATOKTENL-----TPDEVSKREGINARQIVIKITDQGVYTS 62
      277 TQKNDKNIEY--DQCTSSKENIEDNVSRVNDPSPNQKNLNN-----NITDNNIIPSN 329
      63 ---GDHYHYNGKVPYDAIISEELLMDPNYKLDKEDI VNEVKGQYIKVDGKYVYVYK 118
      330 VOIEKOSTLSKKKKKNEKDSYININNSLTNDDQNLKREDIKFNDKAGKITKYD--MLNIK 386
      119 DAHAADNRTKEI-----NRQKQHSQRREGTFRNDCAVVALARQGYTTDDGYI 170
      387 ---NNISIKKIKINDCHSPNENKKNKDNHNO-----CEDNSI 418
      171 FNASDIIEDTGDYIVPHGDHYIIPKNEISASELAALAEAFLSGRGNLSRTRGRNSD 230
      419 -----NICN-----NKN 425
      231 NTSRTNWVPSVSNPQTTN--TNTSN-----NSNTNSQASQNDI-----DSLKQLY 275
      426 NNIGQTNNI--NDNTVNEKINNTSKQKQMLNNTQNNNSEKKYDVVIEQLVNEIDLKKN 481
      276 KLPQSGHYVESDGLV-----PPQAQITSTAAGVAVPHGDHNFIPYSGMSELEERI 327
      482 KQTKKKKQINBPYPYKHLRPSNSDPSLTS-----YSNIHALQETL 523
      328 AR-----IIPLRYSN-----HWVPD--SRPEQSPQPTPPSPQAPAPNLK 368
      524 TKRPHYNYTFLANNEEKRDNQMNRYLHRLPDLCKLIDQDSDTEEDDLSDVDQNK 583
      369 IDSNS---SLVQLVRYKVGEGVFEKGISRYVFAKDLPESETV-----KN 410
      584 QNKQQLLEVVLPPNKKY PNSNDVYKFFEKDINKRPIYCDMNNHLIHEBALRHELYMKKNKA 643
      411 LESKLSKQESVSHLTAKKENVAPRQD-----FYQAYNLTLEAKALFXKNG 459
      644 IDSNNYMDLDGNNQSHKVVYINTEDGEYCIARPYPVSYVYHEKSCYKICDLNLSWIDDS 703
      460 RNSDFOA-----LDKLERN 475
      704 RYAEIOTROYRPAPEYLKSGFNETADINSFACVPELVYTGDFLFPQKDRIDKQKEHLS 763
      476 -----DESTNK-----EKLVDLL 488

```

```

Db      764 FIEVLGNIPKIMIDAGYNSHKYFNKNYRLKININIKKYGLKILKYKNLPEKEISPL 823
Oy      489 LAFLAPIHBERLKGNSQIEYTE-----DEVRIAQLADKYTSDGYIPDEHDISDEG 542
Db      824 CSFLPLMLSVDPQTPSA--YTMLOHPMLNWNVSLEBGDDMYINDESVSINNDRMKNS 880
Oy      543 DAYVTPHNGHSHWIGKDSLSDKEKYAA-----QAYTKEG----- 577
Db      881 NSNNPIYGHNSKKKNKS-SNKKKIDVYKIGNNGNNAVNDVYKYNKNNKNNKPNDD 939
Oy      578 ILPPSPDADVKANPPTGSAALYTRVKEKRIPLVRLPYMVEHYEVKGNLLIPKHQHY 637
Db      940 VPEPSDDYMHANNNNDIVHAV-----LYEKPYNSNNVISYTN-----NKGHK 982
Oy      638 HNIRKAMF-----DHTYKAPNGYTLEDLFAITIKYVEHPDERPHSGDGMNASEHVLGK- 692
Db      963 NNFDFINYQHRRNNNSKNKONISLTNDYTFNSDYIANMMDHDTKQIILKIPAHQISKL 1042
Oy      693 KDHSEDPNKNPKADEPVE 711
Db      1043 KD-----GKNFKAYNESIQ 1056

```

RESULT 26

```

T30290
AAS surface protein - Staphylococcus saprophyticus
C:Species: Staphylococcus saprophyticus
C>Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T30290
R:Hell, W.; Meyer, H.G.W.; Gattermann, S.G.
Mol. Microbiol. 29, 871-881, 1998
A>Title: Cloning of aas, a gene encoding a Staphylococcus saprophyticus surface protein
A:Reference number: Z20809; MUID:98389318; PMID:9723925
A:Accession: T30290
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1463 <HELU>
A:Cross-references: EMBL:AJ000007; NID:e1295629; PID:e1295630; PIDN:CAA03852.1
C:Genetics:
A:Gene: aas

```

```

Query Match      3.7%; Score 154.5; DB 2; Length 1463;
Best Local Similarity 17.9%; Pred. No. 2.1;
Matches 130; Conservative 112; Mismatches 274; Indels 209; Gaps 30;

Oy      13 KENNRSVYIDGKQATQKTENLTPDEVSKREGINAEQIVIKITDQGYVTSKGDHYHYNGK 72
Db      148 EENNNTDSIQSETNSTNEASQTSNDVSSNAEETNTDENSADVANO----- 191
Oy      73 VPYDAIIESEELMKDPNYKLKDEDIV-NEVKGQYIVKDGKYYVYLKDAAHADVNTKEE 131
Db      192 -----NEPAAQNDKAKETSNEDVASSDVK-----QDQTHSD-NAADD 227
Oy      132 INRQKQHSQRREGTFRNDGAVALARSGRYTTDDGYIFNASSDIIEDTGAYIVPHGDH 191
Db      228 LTDQNESETLNDNAVSSNEDVASSDVKQDGTTHSD-----NASDQLTQNES--VAQND- 279
Oy      192 YHYIKNELSASELAALAEFLSGRGLNSRTYRQNSDNTSRITWVSVSGPTTNT 251
Db      280 -----KASTSNEDVASSDVKQDTHSDNASSDVAQNESETQNDN--AESNEDVASSD 332
Oy      252 SNNSTNSQASQNSNDISLKLQYLPISQRHVESDGLVFPDQITSR-TARGAVPHGD 310
Db      333 VNQDTHSDNASSDVK-----DQNESETQND--DKAETSNEDVASSDVKQDD 378
Oy      311 HYHPIPSQMSLEERIRIIPLRYSRNMWVPDSRPEQSPQTPPEPSGPQAPNLKID 370
Db      379 THSDNASSDIADQNESVA-----QNDKAKETSNEDVASSDVKQDTHSD 420
Oy      371 SNNSLVSQLVRKVGEGYFEEKGISRYVPACDLPEYTKNLESKLSKQESVSHLTAKKE 430
Db      421 ANASDIAD-----QNEGATODDKA--TSKED 444

```

```

Oy      431 NVARDOEFDYDKANLITAEAKALFXNKGRRNSFOALDKLERLNDESTNKEKLVDDLAA 490
Db      445 DVNSNDKQDNKAKVNSIIEKASTA--ENKQDPATPSA--KYTPKLRVATTSANTAVATRSA 500
Oy      491 FLADITH---PERLGPNSOI-----EYDEVR-IAQLADKYTSDGYIFD 533
Db      501 VTKEATRALPKYSPKVNSSINNYIRKNNPKAPNYVEDIANYLPQNYNRGRGEQIWM- 559
Oy      534 EHDIIISDEGAYVTPHNGHSHWIGKDSLSDKEKYAAQAYTRKEKILPPSPDADVKAPPTG 593
Db      560 -HDTAND--NSTTIGEINY-----MKNNYT- 581
Oy      594 DSAALYTRVKEKRIPLVRLPYMVEHYEVKGNLLIPKHQHYNIRKAMFDDHTYKAP 653
Db      582 --SAFVAYVYDGRILFTANTDYLAWGAGPQANDRFHVELVHTHDY-----DSFARSI 633
Oy      654 NGYLEDLPAT-IKYVEHPDERPHSGDGMNASEH-----VLGKQHSDEPN-----KN 702
Db      634 NNVA--DYAATNLQYGLVDPDSAEY--DGVGTWTHQAVSNYLGSDHS--DPHGYLEAHN 688
Oy      703 FKADE 707
Db      689 YSYDE 693

```

RESULT 27

```

E71619
RAD2 endonuclease PF0265c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C>Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: E71619
R:Gardner, M.J.; Tetteh, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
Science 282, 1126-1132, 1998
A>Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743; PMID:9804551
A:Accession: E71619
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1516 <GAR>
A:Cross-references: GB:AE001383; GB:AE001362; NID:93845135; PIDN:AACT1842.1; PID:9384513
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PF0265c

```

```

Query Match      3.7%; Score 154.5; DB 2; Length 1516;
Best Local Similarity 17.3%; Pred. No. 2.2;
Matches 165; Conservative 134; Mismatches 326; Indels 329; Gaps 40;

Oy      22 DGKQAT-----QKTENLTPDEVSKREGINAEQIVIKITDQGYVTSKGDHYHYNGKVPYDA 77
Db      262 DGDGTYENINKLDRKKDENNLSSINYN---KVQD--VNNND----- 301
Oy      78 IISEELMKDPNYKLKDEDIVNEVKGQYIVKDGKYYVYLKDAAHADVNTKEEINR-- 134
Db      302 -----DDKDXKXENINEVYR-----DQKNVYV-----KKKENINNYL 334
Oy      135 ---OKQHSQRREGTFRNDGAVALARSGRYTTDDGYIFNASSDIIEDTGAYIVPHGDH 191
Db      335 DDDDEKEDIONKNG-----VYNNDDIDEQIIRKKMMARKKY 370
Oy      192 YHYIPKA-----ELASAS-----SRTYRQNSDNTSR----- 203
Db      371 YESIPKTFKGLCMRRPVDIILISNYTMELEISETIKVHENKFKQHLNVLDENNSTPVPV 430
Oy      204 -----ELAAEAFLSGRGLSN-----SRTYRQNSDNTSR----- 234
Db      431 NMNLLKNINXKQNDLIEGEEKKSFL---NLIVDSCYSSNSRLENDEIERGKIMF 486
Oy      235 -TNVPSVSNDETTNTSNNSTNSQASQNSNDI-----DSLKLQYLPISQGHVE 285
Db      487 ITNDEKISININNNNNNNNNNNNDNDVILIEHNKNNMNIYDNKYNVVECSSEKIN 546

```



```

Db      3236 -----RGDDSPDSS-----PEEQSVIE-----IPTAPMENVPFTEKSKIPVR 3274
QY      351  PQTPEPSPGPAPNLKIDNSSLSVQVRKVGSGY---VPEEKGISRYVAKLPSBT 407
Db      3275  TWPSTPAP-----PSAEYESS-----VSEDFLSSVDEENKADAKPKSLPKV 3319
QY      408  -VKNLESKSKOESVHTLTAKKENAVAPRDOEFYKAVYLLEAKKALFXNKGRR-SDFO 465
Db      3320  PLQREEQQLSDLD-----TSVQKTVAPQGDMAIAPD-----NRSKESDAS 3362
QY      466  ALD-----KILRLNDESTNKEKLVDDLFLAPDI-----495
Db      3363  SLDSKTKCPVKTRSYTETETESRERAELELESEEGATPKILTSLRL-----PVKSRSTTS 3418
QY      496  -----THERLGKRN-----SQIYTE---DEVNLAQADKYTTSDDGIPEHDIIDE 541
Db      3419  SCRGSTSPKESKEHFPDLVYRNSIEFPFEISDEA--SKVLDRLTQSE---REQIVSDD 3472
QY      542  GDAAVYTPHMGSHWIGKDSLSDKEKVAQAAYTEKGI.LPPSPDADVKANPTGDSAAIYN 601
Db      3473  -----ESSAL--3478
QY      602  RVYKGRKIPVRLPYMVEHTVEVKNGNLIIPKDHVHNIKFAWFDHTYKADNGYTLDEL 661
Db      3479  EVSVIENLP---PVETEHSVP-----BDI 3499
QY      662  FAT-----IKYVVEH-PDERPHSNDGNGNASEHVLGKKHSEDPNKNFKADEPVEET 713
Db      3500  FDTRPIMDESLETLRIEIPDENGH-----DIAEDP---QDEORIEER 3539
QY      714  PA-----EPYQVETEKEVAOLKEAEVL-----AKYDSS 745
Db      3540  LAYVDHLGFSWTELARELDFTEBOIHQIRIENRPSLDQOSYLLKIMLERGKATITGN 3599
QY      746  L-----KANA-----TETL-----AGLRNLTLOIMNNSIMAEKTLALL 782
Db      3600  LVECLTKINRMDIVHLMETNETPELQERISHSYAEIEQITTLDSSEGSFVLQF-ELCTAQH 3658
QY      783  KGSNSSVSKE 793
Db      3659  KOKEQAVSKE 3669

RESULT 29
S23467
Probable long-chain-fatty-acid-CoA ligase (EC 6.2.1.3) - malaria parasite (Plasmodium fa
C:Species: Plasmodium falciparum
C:Date: 22-Nov-1993 #sequence_revision 01-Dec-1995 #text_change 03-Jun-2002
C:Accession: S23467
R:Zhao, Y.; Kappes, B.; Yang, J.; Franklin, R.M.
Eur. J. Biochem. 207, 305-313, 1992
A:Title: Molecular cloning, stage-specific expression and cellular distribution of a put
A:Reference number: S23466; MUID:92331669; PMID:1378403
A:Accession: S23467
A:Status: preliminary; translation not shown
A:Molecule type: mRNA
A:Residues: 1-749 <ZHA>
A:Cross-references: EMBL:X63648; NID:99937; PID:99939
C:Superfamily: human long-chain-fatty-acid-CoA ligase; acetate-CoA ligase homology
C:Keywords: acid-thiol ligase; coenzyme A
F:103-732/Domain: acetate-CoA ligase homology <ACL>

Query Match 3.7%; Score 154; DB 2; Length 749;
Best Local Similarity 18.7%; Pred. No. 0.84;
Matches 114; Conservative 92; Mismatches 209; Indels 196; Gaps 26;

QY      168  GY-IRNADIEDTGDAYIVP-----HGDHYHYIPKNELSASELAABAFISGRGNLSN 220
Db      201  GYKLYYYKDLIDKYKRNIIIPOTEHHDHDFKEHPNNAF-YDKLKEQADYNNNNNNNN 259
QY      221  SRYRRONSDNTSRTNWVPSVNPQTNTNTNSNTNSOASQSDND-IDN-----LLKQL 274
Db      260  NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 308

```

```

QY      275  YKLPLOSRRHVESDGLVFPQAQITSRTAGVAVPHGDHYHFI-----315
Db      309  KAKGNANDVYT--IIFISG--SSGTRGVWITNSFTFLQALIDGNRLGLKKYDVVF 364
QY      316  PYSQMSLEBEIARI1-----PLVRSNH---WVPSRPQSPQPTPEPSPGPAPN 366
Db      365  SYPLAHYERFIEYAVCFGHIKIGFSGNIKELVGDNNELKPFLLT-----412
QY      367  LKIDNSSLSVQVRKVGEGVFEKGI-SRYVAKLPSRTVKNLESKLSKQESVHTL 425
Db      413  -----VRLIQKHDSVM---EGIKSXFIARLVKTSLKNKTSAYVK-----452
QY      426  TAKKENAVAPRDOEFYDAVYNLLTEAHKALFXNKNRNSDQALDKLRLNDESTNKEKLV 485
Db      453  -----NSRKHSHKIYDLIPIKNNKGGCIKT-----QVMSSSNDKXKLI 493
QY      486  DLLAFLAPITN-----PERLGKRNQSIQIYTEDEV-----515
Db      494  DLQWIFSPPISEGMGTEVGVLQHRFDSFTKGTIGGFANVEFKVYVQWVKYDPRKYP 553
QY      516  -----RIQLADKYTTSDDGIIDEHDIISDEGDAYTTPHMGSHWIGK 558
Db      554  NKGEICVKGSSLMVGYFRDELTTRKCFDEDFPLTGDVVEVNNNAV-----601
QY      559  DLSLDEK---KVAQAAYTEKGI.LPPSPDADVKANPTGDSAAIYNR---VKG---EKR 608
Db      602  KITDRAKNIKLAQGEY-----IPEKLENLSN-----SIYENITVHGYSYNE 647
QY      609  IPLVRLPYMVEHTVEVKNGNLIIPKDHV-HNIKFAWFDHTYKAPNGYTLDELFAPIKY 667
Db      648  LVSWVLPRQIMLSYAKKHNLLNPYEDLLKHDIIIKLFOEILNISKYKLNGL---EKI 704
QY      668  YVEHPDERPHS 678
Db      705  HLFHLTSTPFS 715

RESULT 30
A36685
205K microtubule-associated protein - fruit fly (Drosophila sp.)
C:Species: Drosophila sp.
C:Date: 12-Apr-1991 #sequence_revision 12-Apr-1991 #text_change 24-Sep-1998
C:Accession: A36685
R:Riminger-Finger, I.; Laymon, R.A.; Goldstein, L.S.B.
J. Cell Biol. 111, 2563-2572, 1990
A:Title: Analysis of the primary sequence and microtubule-binding region of the Drosophi
A:Reference number: A36685; MUID:91115949; PMID:1703540
A:Accession: A36685
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1163 <IRM>
A:Cross-references: GB:X54061; NID:98201; PID:98202
C:Genetics:
A:Gene: FlyBase:Map205
A:Cross-references: FlyBase:FBgn0002645
C:Keywords: alternative splicing

Query Match 3.6%; Score 151.5; DB 2; Length 1163;
Best Local Similarity 20.3%; Pred. No. 2.2;
Matches 160; Conservative 105; Mismatches 310; Indels 215; Gaps 35;

QY      122  HADVNRKKEI-MNOKQ-----HSHQREGCTRRNGCAVALARSQGRYTTDDGIYF 171
Db      4  HEDNAQLDNYLQNRLASLQICGAGSHNPPLADATGNGCAPGIAPSK-----52
QY      172  NASDIIEDTGDAYIVPHGDHYHYIPKNELSASELAABAFISGRGNLSN 221
Db      53  --SDEVGEED-----BEWKYI--HEVROSEKLOQEKLPJLTKETNGFPQGR-----95
QY      222  RYRRONSDNTSRTNW-----VPSVNPQTNTNTNSNTNS---OASQSDND 266
Db      96  -----DSDQVHGNGAAAVFNLVEEDVEVYKNDGDESTNSNTTSTTSDVEVARQAOEPNQ 149

```


Db 1199 SE---KKNKEIKNEQIRIIDEVAKNNKNAKILDI--QLSVEPKIKFLKIDLRKTS 1251
 QY 354 ---TPESPQPOPAPIKIDS-----NSSLSQLVKRGEGYVEEKGISRYVPAQDL 403
 Db 1252 DDCLEKTOIERKISLSIDTQETKLIENKNIANTL-EKLBSLKQKKNIEQKQELDE 1310
 QY 404 PSETVKNLESKLSK-QESVSHLTAKKENVAPRDEFFYKAYNLLTEAHKALEFXNKRNS 462
 Db 1311 VNSKINIESNVQHKKNYIEGIVEKINEIAKANKQIESTQKLIPTIKNLI-SFKAN 1369
 QY 463 DQOALD--KLLEBNDESSTNKEVLVDLAFAPITHPERLGRKPNQIETVE---DEVR 516
 Db 1370 DLEGIDTNKNGKGYNEMNN---IYEEFKSYDLTH-----YLEVSKPEPT 1414
 QY 517 IAQLADKYTTSDGYIFDE-----HDIISDEGDAYVTPMGHSHWIGKDSIDKE 565
 Db 1415 YEOIKKKRITIAQNEELLTNKNNKAKSYLDDIEANEDRIVT-----HF--KKNKLDVA 1466
 QY 566 KVAQAQYTKKGLIPSPADVKNAPTGSAAAIYNRVKGKRIPLVRLPYWEHTVEVK 625
 Db 1467 DKFTNEYS-----KVNKGFDNISINSINNVK-----KSTDEML 1498
 QY 626 NGULLIPKDHYNINIKFAMFDDHTYKAPNGYTLIEDLPAITKYVEHPDERPHSNDG--W 682
 Db 1499 LANINQOTKEMVANIYSKKY--YSYK---YEAENIFINIPKLANSINLOIKSSSGIDLF 1552
 QY 683 GNASEHVLGKDESEDPNKNFKADEEVEET-----PAEPVPOVEETKVEAQLKRAEVL 737
 Db 1553 KNININILPYLDQCKDQTLTFIPSPKSTETTYKIDSINTLADILKRSOELOKQKQOAL 1612
 QY 738 LAKVTDSSJ--KANAT---ETLAGLRNNLLQIMDN-----NSIMAEAEKLA 780
 Db 1613 NIIFERRLHDKVQATNELKDTLSLKNKKE-QILNKKVLLHKSNELNKLKSCNSQNYDT 1671
 QY 781 LKGSNPPSSVSKKIN 796
 Db 1672 ILESKYDKI-KEKSN 1686

RESULT 34

protein-tyrosine-phosphatase (EC 3.1.3.48) 2E - rat
 C.Species: Rattus norvegicus (Norway rat)
 C.Date: 10-Apr-1996 #sequence_revision 03-May-1996 #ext_change 22-Jun-1999
 C.Accession: S51005; S51161
 R.Li, Abbe, D.; Banville, D.; Tong, Y.; Stocco, R.; Masson, S.; Ma, S.; Fantus, G.; Shen, FBS Lett. 356, 351-356, 1994
 A.Title: Identification of a novel protein tyrosine phosphatase with sequence homology to A.Reference number: S51005; WUID:95104449; PMID:7805871
 A.Accession: S51005
 A.Molecule type: mRNA
 A.Residues: 1-1175 <LAA>
 A.Cross-references: EMBL:U1971; NID:g662113; PIDN:AAA62153.1; PID:g602255
 A.Genetics: PTP2E
 A.Accession: S51161
 A.Molecule type: mRNA
 A.Residues: 840-1175 <LAB>
 A.Cross-references: EMBL:U18293; NID:g603228; PIDN:AAA62154.1; PID:g603229
 A.Genetics: PTP2E1
 C.Genetics: <PTP2E>
 A.Note: clone PTP2E
 C.Genetics: <PTP2E1>
 A.Note: clone PTP2E1
 C.Superfamily: protein-tyrosine-phosphatase, nonreceptor type 14; protein 4.1 membrane-b C.Keywords: phosphoprotein; phosphoric monoester hydrolase; tyrosine-specific phosphatas F;25-304/Domain: protein 4.1 membrane-binding domain homology <B41>
 F;922-1157/Domain: protein-tyrosine-phosphatase homology <PTP2>
 F;1109/Active site: Cys (phosphocysteine intermediate) #status Predicted
 F;1115/Binding site: substrate phosphate (Arg) #status predicted

Query Match 3.6%; Score 149.5; DB 2; Length 1175;
 Best Local Similarity 19.4%; Pred. No. 2.9;
 Matches 138; Conservative 87; Mismatches 273; Indels 213; Gaps 30;

QY 29 KTNLTPEDEVSKEGINAEOIVIKITDQGYTS-----HGDHYHYNGKVPY----- 75
 Db 202 EAEMLYMEVEEMDGYGEESYPAK--DSQSDSISIGALDGIIVKIKNGRPVYFRMHDA 260
 QY 76 ----DAIISEELMKDPNYKLDKEDIVNEVKGYYIKVGYVYLKDAAHADNVRTKE 130
 Db 261 NMSHNKSPFALTELAKKEETIQOTEDMET-----AKYVRLCYARHAKFRLNQC 309
 QY 131 EINROKQEHSCHEGCTRRNDGVALASQ-----GRYTDGGIIFNASDI 177
 Db 310 NLQTAATLNSVRRGSSSR---MSLPKQPYAMPPPQLHYNGHY-----EPF 355
 QY 178 EDTGAYIVPHGDHNYIPKELASAEILAAEAFISGRGNLSNRTY--RRONSNTSRT 235
 Db 356 ASSQDNVFPFNKNGCYHCSQISLDRTOID-----LSGR--IRNGSVYASHNSINTQOP 408
 QY 236 NNVPSS--VSNPGTNTNTSNSTNSQASQ-----SNDISLKLQLYK-LPLSORH 283
 Db 409 YLOPSPMSSNPSPISPDVWRPDYIPSHHSALIPSPRPTPDYESVMKRLNGVHADRH 468
 QY 284 VES-----DGLYFDPQAQTSRTARGVAPHGDNHF-----IPISQ 319
 Db 469 SHSLNNLIGSSVAYSAPDALVYQPEI--REHPILASPOSANHYFPNLYSFHSQAPYV 526
 QY 320 MSELERLARIIPLRYSNHWVPDSRPE-----QSPQPT-----PESPGPO 362
 Db 527 PYERRPVYAV-----SVPLTNVQLOADYPAINRTQVYRRPPPPYPR 573
 QY 363 PA---PNLK-----IDNSSLSQLVKRGEGYVEEKGISRYVPAQDLPSETVKNLESK 414
 Db 574 PANSPTDLSRLHYISSNPDILTRRVHNSVQ--TQOESLPVAHSLQGESEPLTARH 631
 QY 415 LSKQESV-----SHTLTAKKENVAR-----DOEFYKAYNLLTEAH- 451
 Db 632 LQKRNSEIYAGLTGHEGRLKEETMSASADVAPRTSAGSQSVFSFKVQOEGTEEG 691
 QY 452 KALFYNKGKNSDPQAL--DKLLERLNPESTKEKLVDDLL-----AFLAP 494
 Db 692 SCGYHKKSLSATMLIHSESEDEDDSSFEHAVSERLTAAASQEQQLYPCASVTP 751
 QY 495 ITHPERLGNPSQIEYTEDDEVRIAQLADKYTTSDGYIFDEHDIISDEGDAYVTPMNGSH 554
 Db 752 VGPLHIFEPKSHV--TEPEKAKADISPVHLM-----THQPRRH 790
 QY 555 WIGKQSLSDKEVVAQAQYTKKGLIPSPADVKNAPTGSAAAIYNRVKG 605
 Db 791 GLTPSMSESDDLTTSGRYARADSL-----KKRPVSDLLSGKKNTEVG 833

RESULT 35

hypothetical protein SA1562 [imported] - Staphylococcus aureus (strain N315)
 C.Species: Staphylococcus aureus
 C.Date: 10-May-2001 #sequence_revision 10-May-2001 #ext_change 22-Oct-2001
 C.Accession: A89959
 R.Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc ma, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogaewara, N.; Hayashi, H.; Hatamatsu, K. Lancet 357, 1225-1240, 2001
 A.Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus. A.Reference number: A89758; WUID:21311952; PMID:11418146
 A.Accession: A89959
 A.Status: Preliminary
 A.Molecule type: DNA
 A.Residues: 1-1274 <KUR>
 A.Cross-references: GB:BA000018; PID:g13701536; PIDN:BA842830.1; GSPDB:GN00149
 A.Experimental source: strain N315
 C.Genetics:
 A.Gene: SA1562

Query Match 3.6%; Score 149; DB 2; Length 1274;
 Best Local Similarity 20.1%; Pred. No. 3.4;

Matches 153; Conservative 119; Mismatches 273; Indels 216; Gaps 39;

```

OY 111 GKYVYVYLDAAHADV-RTKEINROKQESHQREGGTRNGAVALARSGQRYTTDDGY 169
DB 52 GKFRRPMVAAYENEVQSDADTISDEKEY--HRYRKQSHD-----SRSQRRHRRRNQ 104
OY 170 IFNASDIEDTDGAYI---VPHGDHYH---IPKNELASASELAALAEAFLSGRNL-SNS 221
DB 105 TTEEQNYSEQRGNSKISQOSIKYKQSHHYHTNKPCTGYSAINGIEKTHKSTHMYSN 164
OY 222 RYTRONS-----DNTSRTNVPSVSNPGTNTNTNNSNTNSQASQSDIDSLKQLYK 276
DB 165 TMRHAKDSTPDYHKESFKTSEVPSPAF-GTMRPKLENG-----R 203
OY 277 LPLSQ--RHVESDGLVPDPAQITSRARGVAVPHGDHYHFIYSQMSLEERLARIPLR 334
DB 204 LPVSPKSEKESDOKKYDYVAKTQT-----SONKOLEQ----- 238
OY 335 YRSNHWVPSRPEQSPPTPEP--SPGPAPNL-KIDNSGL-----VSQLVKVEGY 387
DB 239 -KQ-----DSVVKQGTASQSDENVSSTTKSMPTSKYDNTIKENIYASQIVEIRER 293
OY 388 VFEKGIKRYVFAKOLPSETVKNLESKLSQESVSHTLTAKKENVAPR-----DOE 438
DB 294 --ERKVLQKRRKKAQQ--KREHKNEQDAIQRAIDEMAKQAERYVGDSDLDDSD 348
OY 439 FYDKAVNLLTEAHKALFXNKGNSDFOALDKLERLNDESTNKEKLVDDLAFLAPITHP 498
DB 349 LTDNS-----TDSQ--LHTNGIENETVSDENKQASIQEDTN-DTHVE--SPNYE 397
OY 499 E-RLKPNQSQLEYTEDEVARIAQLADKYTTSQGYIFDEHDI-ISDEGDAVVTPEHMSHWI 556
DB 398 EVSLNQVSTTKQLSDDEVTVSNVTSQHQA-----LQNHVEVD----- 436
OY 557 GKDSLSDKEKVAQ-----AYTEKGIPLPSPDAD-VKANPT--GDSAAIYNRVKEK 607
DB 437 -KDELKQNSRLIADSEEDGATKEEYSGQIDDAFEYELNDEVEDTTSNIEDTNNNA 495
OY 608 RPLVRLPYMVEHTVEVKNGLIIPHKHYNIKRA--WPDHYYKAPNGVTTLEDLPA 664
DB 496 SEMHDAKKTQEHATVESQVNNIDKTV--NEIELAPHHKKDDQNLNVLNSLKTNDV-- 550
OY 665 IKYVHEHDERPHSNMGNASEHYLKG-----KDHSEDPKQNKAD---EEPEETP 714
DB 551 -----NDG-----HVEDSSMMEIKHNAETENQNAABEQVNEBKT 590
OY 715 AEPEVQVETEKVEAQLKEA-EVLL-----AKVTDSSLK----- 747
DB 591 IENVAPKQTEKYSTLSKRPFNVMVTPSDKKRMMDRKKHGVNVDELKPVQSKQAVSESK 650
OY 748 -----ANATELAGLRNLLTQIMDNNSIMAEAE 776
DB 651 PASQAPSSRSDSQESNTNAYKTNMTNNTNNVNNOLIGHAE 691

```

RESULT 36

reticulocyte-binding protein 1 - Plasmodium vivax
 C:Species: Plasmodium vivax
 C:Date: 04-Mar-1993 #revision 18-Nov-1994 #text_change 28-Apr-1995
 C:Accession: A42771
 R:Galinski, M.R.; Medina, C.C.; Ingravallo, P.; Barnwell, J.W.
 Cell 69, 1213-1226, 1992
 A:Title: A reticulocyte-binding protein complex of Plasmodium vivax merozoites.
 A:Reference number: A42771; MUID:9215338; PMID:1617731
 A:Accession: A42771
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-2829 <GAL>
 A:Experimental source: Belam strain, merozoites
 A>Note: sequence extracted from NCBI Backbone (NCBIN:108114, NCBI:P:108115)

Query Match 3.6%; Score 149; DB 2; Length 2829;

Best Local Similarity 17.2%; Pred. No. 11;
 Matches 166; Conservative 148; Mismatches 336; Indels 314; Gaps 42;

```

OY 24 KQATQKENTLTPDEVSKREGINAEQ-----IVKITD-----OGY--VTS 62
DB 1662 KESYEKLETVQNMBSR---INVEGSLTIDDKITITENDLKKKQYEGGLQOKIKEN 1718
OY 63 GDHYHYNGKVPYAIISEELMKDPN-----YKLEDEIVNEKGVYKVGDKYVY 116
DB 1719 AD-----KRSNFELVGSSEIALLDPSTSFIFIKLKEIYDTGOLK-NYGVKMEIHGEF 1772
OY 117 LKD-----AAHADN-----VTKBEINROKQESHQRE----- 144
DB 1773 TKSYNLLETHTLSNATDYSVTFEKAQSLRELAKEKEEHLRREBEAFLINDIKKVESLKL 1832
OY 145 -----GTPNDGAVALARSGQRTTDD-----GYFASDIIETDG 181
DB 1833 LKEMMKVSAEYEGMKDHTSVSQLVQDMKTIIVELKTLNDISECSSVLNNVSAIVKKV 1892
OY 182 DAYIVPHGDHYHFIIPKNELASASELAALAEFLSGRNLNSRTYRONSNTS---RTWV 238
DB 1893 ES---KADYR--DANSMTESWTLANYFLSDAKISGMEFNAEMKSNFKTLELEIF 1947
OY 239 PSVSNPGTNTNTNNSN-----TNSQASQSDIDSLKQLYKLPQS RHVESDGLV 290
DB 1948 SVISNSNELLKKIEQSDNDVYQKERESQOLAKQADTYNYIK--LKQEFNEKLEAKNKE 2005
OY 291 FDPQOITSRTARGVAVPHGDHYHFIYSQ--SELEERLARIPLRYRSHHWVPSRPE 347
DB 2006 EVSEKRYEALKRISQVEGIRCHFPENPHRLDNTBELENKQVTVI-YR-----DKSE 2058
OY 348 QSPQPTPEPSPGPAPNLKIDNSGLVSQLVKVEGYFEEKGISRYVFAKDLPSET 407
DB 2059 RES-----GLQEMEN-EMNTYSNITOL-----EGIVSAGSEKDEIEKLERSNE 2103
OY 408 VKNLESKLSKQESVSHTLTAKKENVAPRDOEFYDKAVNLLTEAHKALFXNKGNSDFOAL 467
DB 2104 MRNISEKIS-----TIDSKVIEMNSTIDELYLKGN--COAHMISLISYTA--MKT 2151
OY 468 DKLERLNDESTNKEKLVDDLAFLAPITHEERLKRPSQLEYTEDEVRIQLADKYTTS 527
DB 2152 SKKLIMINKEKENTKCV-----YIKDSSST 2179
OY 528 DGYI-----FDEHDIISDEGDAYVTPEHMGSHWIGKDSLDEKVAQAAYTKE 575
DB 2180 DGYVETLKGFTYSGKLTFTSSASEIYQNDTYSVNFABE---KESLVAIRIKKELYLPH 2235
OY 576 KGIPLPSPDADVKANPTG--DSAAAIYNRVKEK--IPLVRLPYMVEHTVEV-- 624
DB 2236 Q-----NSDISIVEGGVQNMALDYDKLNEKKEEMDELYRNISETKLKQ--EHSTDVFK 2287
OY 625 -----KGNLIIP----- 632
DB 2288 PMIELHKGMEITNNSKLSLEKEKKLSVNDHMSMEAMIKNGLYKTPESVONINNYSVI 2347
OY 633 -----HKDHY-----HNIFEAWPDHT-----YKAPNGVTTLEDLPA 665
DB 2348 EAEVKTLEIDRQDGYVQYVEEHKQPSILIDNTNALMDIEIFKKNYNL----- 2400
OY 666 KYVHEHDERPHS-NDMGNASEHYLKGKQHSDEDPNKNFPADEBPVEETPAEP----- 717
DB 2401 ---MEVNTETIHRANDYIEKTNKLVQAKTEVEBQILINIKQNDMLQNIPLKVSIIIEYF 2457
OY 718 EVPOVETEKVEAQLKEAEVLAKTDD--SSKANATETLAGLRNLLTQIMDNNSIMAEAE 775
DB 2458 ENVKKKESLINDLYEOERLL-KIGEHLDIKRNVETLSSYEIDQKEMWSKN-LLEKK 2515
OY 776 EKLL 779
DB 2516 SKQM 2519

```

RESULT 37


```

Db 361 H-----VSEVTAAKNAATELNQMGLEQAHIDQNTVKGQVFTDADAKRD----- 407
Qy 442 KAY-NLLEBAKALFXNKGKNSDFQALDKLER-----LNDES--TNKEKLVDDLLAF 491
Db 408 -AYTAVAARAEITLTKTGANTSKODVEALITONVTSARNAALNDQNTNNAKAAGNALNN 466
Qy 492 LAPITHPRRLGKPN-----SQIEYTEDEVRIAQ-----LADKYTT--SDGYIF 532
Db 467 LTIINNAQRDLTTYIIDQATTYAGVAVSNTGTQNTAMANIQNGINDANTLASENY-- 524
Qy 533 DEHDIIISDEGAYVTTPHMGSHMICKSLSDKEKYAAQ-----AYTKEGILPPS--PDA 585
Db 525 --HDADSKKTATQAVTNNAENILNKNGSNNLDKAIVENALSQVTNAGALNGNHNLDQA 582
Qy 586 DVKANPTGDSAAATYNRKVGEKRIPLVLPVVEHTVEYKNGNLIIPKHDIHNIKFMF 645
Db 583 KSNANTT-----INGLOHLTTAOKDKLKQOVQOQON-----VAGV 617
Qy 646 DDHTYKANGYTLLEDLPATIKYVE-----HDERPHSNDGNGNASEHVLGKKDH 695
Db 618 D--TVKS--SANTLNGAMGTLRNSIDONTATKNGQNYLDATERKKTNNNAVDANGVINA 674
Qy 696 SEDPKNKFKR-----DEEPEETPAEPEVPOVETEKVEAOLK-----EAEVLLAKV 741
Db 675 TSNPMDANAIQNTQVYSTKNALDGHNLQAKQTATNLDGATNINKAKQXALKAQV 734
Qy 742 TDSLKANATETLACLNNLT-----LQ--IMDNN-----INAE 774
Db 735 TSAQRVANVT--SIQQTANELTAMQLOHGIDENATKOTOKYRDAEOSKKTAYDAQAVA 793
Qy 775 AEKLLALKGNSPSSVSKEK 794
Db 794 AKAILNKOTGNSDKAAVDR 813

```

```

RESULT 39
T21558
hypothetical protein F29G6.3b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T21558
R:Haris, B.
submitted to the EMBL Data Library, August 1996
A:Reference number: Z19441
A:Accession: T21538
A:Status: Preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1785 <MIL>
A:Cross-references: EMBL:Z78543; PIDDN:CAB01754.1; GSPDB:GN00028; CESP:F29G6.3b
C:Genetics:
A:Gene: CESP:F29G6.3b
A:Map position: X
A:Introns: 10/1; 43/3; 61/3; 88/3; 110/3; 144/2; 183/3; 218/3; 252/3; 288/3; 1083/3

```

```

Query Match 3.6%; Score 148.5; DB 2; Length 1785;
Best Local Similarity 17.3%; Pred. No. 5.9;
Matches 134; Conservative 126; Mismatches 305; Indels 211; Gaps 37;

Qy 12 VKENNRVAYIDGKQATOKTENLTPEV-----SKREGINAQIVIKITDQGVTSHGSHY 66
Db 661 IAQSPQVOTIDYFKKINRAHNFVSTPSPLENNKESFHHNSLEAEKRSKSVERA 720
Qy 67 HYNKGKLVYDAIIEBELMKDPNYKLKDE-DIVNEVGKGYIKYDGKYVYVLK----- 118
Db 721 HHY-----LPPSRANVQDKRIIRE-----HHYVLEKKSQSLD 754
Qy 119 -DAHAAD-----VATKEINRQKEHQSHREGGPRPDGAVALARSGRITTDGTYFN 172
Db 755 LKVNHPEHNVYVLYKKTEDHQKVNHEHNFPAEKTEA----- 794
Qy 173 ASDIIEDTGDAIYVH--GDH-----YHYIP-KNE--LSASEL-----AAA 208

```

```

Db 795 VKDVVHEHNYIPLAHTKDEHNANVTREHNYIPLKTEQVOGKTEIERHQVPMTHKET 854
Qy 209 EAFISGR-----GNLSNRT--YRRON-----SNTSKTNVPSVSNPCT 246
Db 855 KSPDGRSHSYIPLGAVNEOSKTEI PROHNVPVVDKKEKKNVEINOGNVVPASHNAV 914
Qy 247 TINTNNSN-----SNTNSQASQNDISLKLQVLYKLPSQHRHVESDGLVFPRA----- 294
Db 915 SDEKPISEHNFVPLATTYETREHNSD-----LSHHNMPYPAKREBKTAAGDKVGEH 970
Qy 295 --OITSRTAGVAVPHGDHYHFIYQSOMSELEERAIIPLYRYSNNHVPDSRPEQSP 351
Db 971 HUYPLASNTER-VGVHGOEHVYPL--IAKPTESHASKIYIQHQ--YIPLAHRHQSV 1024
Qy 352 OPTPEPSPG--OPAPRLKID--SNSLSVQLYKVGEGYVEEKIGSRIFYPADLP--- 404
Db 1025 RLPEHNYIPIAHDHSHQKVDLTNHQVLPPIVNA-----EQKEAVKLYEHQVPLRP 1077
Qy 405 -SETVKLESKLKSQESVSHTLTAKKENVAPRDEQKAVNLLTEAHKALFXNKGKNSD 463
Db 1078 KKDSPANYSSAPOLPREHNYVPAPVRSVEHGHQVHEQ-HNYIPAVH-----HHETHQ 1131
Qy 464 FQALDKLERLNDESTNKEKLVDDLFLAFLPITPERLKGPNQIEYTEDEVRIAGLADK 523
Db 1132 LDRQGYIPIVAAKHEAKKNVLYEH--QYIPITTEQNGVDRRHNY-----VPVGKTA-K 1184
Qy 524 YTSDDGYIFDEHDIIISDEGAYVTTPHMGSH-WIGKQSLSDKEKYAAQATYKEGILPPS 582
Db 1185 ATTTDIIHVOHQVPPQASPKRTISEVRSNHYI--PALPEKHQSRHEHQSHTYPPS 1242
Qy 583 PDADVKANPTGDSAAATYNRKVGEKRIPLVLPVVEHTVEYKNGNLIIPKHDIHNIKF 642
Db 1243 QKSTSHQSP-----PKLIHNSFVPO-----IPKTKDNOV-- 1272
Qy 643 AMPDDHTYKAPNGYTLLEDLPATIKY-----YVEHDERPHSNDGNGNASEH 688
Db 1273 --VNEHGVPP--VKSDDLKTSVDHSKHEVLPVSVHTHESKODFGHSKKGVPVSLH 1324

```

```

RESULT 40
AG1085
hypothetical protein lmo0086 [imported] - Listeria monocytogenes (strain EGD-e)
C:Species: Listeria monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AG1085
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bioecker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurgey, O.; Entlian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karet, U.
Science 294, 849-852, 2001
A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurapk, G.; Madeno, E.; Maitournam, A.; Mai
ok, C.; Schlueter, T.; Simoes, N.; Tlerriz, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A.; Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; PMID:21537279; PMID:11679669
A:Accession: AG1085
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-1959 <GLA>
A:Cross-references: GB:NC_003210; PIDDN:CAC98301.1; PID:gl6409445; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo0086

```

```

Query Match 3.6%; Score 148.5; DB 2; Length 1959;
Best Local Similarity 17.7%; Pred. No. 6.8;
Matches 164; Conservative 125; Mismatches 307; Indels 331; Gaps 39;

Qy 77 AIISEELMKDPNYKLKDEDIVNEVGKGYI-KYDGY----- 113
Db 18 AIVSAFPLMRDADAIEEGTTPSPKGVYDFSEAKYNDTWINNEKTTIDKKEKMPNL 77
Qy 114 --VYLLKD-----AAHADNVRTKEINRQKEHQSHREGGPRPDGAVALA 157

```

```

Db      78 ARTLFLLKGRIOFLGKSAVAVKASADLIELPSTKSTITESKGSYK-----AKSDSNKTLA 130
Qy      158 R-----SQGRY-TTDDGYIFNASDI-----IEDTG- 181
Db      131 QLPKGTIVVLAAGRIYIIDNAYLNKKQGLNKLPRKVLVSIIDENKKMLMGKTIIEELG 190
Qy      182 -DAYIVPHGDHYHYIPKNELSASELAALAEPLSGRNLNSRTYRQNSDNTSRTNWPS 240
Db      191 DDAYIEMENNHYHFDLKKEMLVISQTAKEITDI-----DIRSIKVEIDNAEKR-----S 238
Qy      241 VSNPGTNTNTSNSTNSQASQNSDIDSLKQLYKLPLSQRHVESDGLVFPQAQITST 300
Db      239 IKTAKETEKEITKDSIEKSEKQNSDASQGS-----AQNNESESQDQOTODASGNN 291
Qy      301 ARGVAIVPHGDHYHYIPYSQMSLEERIRIIPLRYSNHWVPDSPEQSPQPTPEPSPG 360
Db      232 TNGSA-----NNNTAR-----SNGTGSNSDNTGSGP-----CG 320
Qy      361 POPAPNLKIDSNSLSVSQLVRKVGEG-----YVFEKQISRYPFAKDL-----PSET 407
Db      321 TQNKN-----EGDVNKANELIKKLEANEANTNTFOVPIVDVNLTVKQGVAAAKLKLIDSQR 377
Qy      408 VKNLESKLSKQESVSHITLAKKENVAPRQEF-----YDKAYNLTTEAHKALFYNNKEN 461
Db      378 LNSLEAIL--YDSKNIVYKKEKLNSTKANQNTFPNNLKXGETYQVAVQGSYKASANKOD 435
Qy      462 SDF--QALD-----KLERLNDESTNKEKLVDDLLAFIAPITHEPRLG----- 502
Db      436 TIFRFQYEAKEVULTPKTERGED-----YMKAEITATELYGHIIDKLVKI 482
Qy      503 KENSOIEYTEDEVRI--AQLADKYTSDG-----YIFDEHDI-----SD 540
Db      483 KENNSNVITISQKVTVDAQSL-----TKDQGVKFNLSLSDKEYIIEEMELLVDGKNVTD 537
Qy      541 EGDVAIVTHMGSHHITGKDSLDEKVAQAQYTKEGILPSP----- 583
Db      538 DGMVVISSTLAKKPTIEGLNLS-----YSTKGGEFVAPVNLVDRBSITSIRYVA 588
Qy      584 --DAVVKANPTGDSAAAIYNRVYKGEKRIPLVRLPVVHEHTVEVKNGLIIPKDHVHNK 641
Db      569 YLEDYKYNGSNAKEYA--YSVVDANOKTAIVK---VGRITVMDNDQNYIFVIGISGNQ 643
Qy      642 FAMFDHTYKAPNGYTL-----EDLFATIKYVV-----EHP 672
Db      644 ---SDYFATPATASNVVVGKTKPTVERSLKEAEDQKLTINVEVFDADNTLILFNLHP 699
Qy      673 DERPHSNDGWNASEHVLGKQD-----HSEDPNNKF 703
Db      700 TLKLYKSDAQGWYSGNPVATVLDLTKKSDITNLLFEDGLSESAVYVVMVMTGSYNLDGAGI 759
Qy      704 KADE-----BPVETPAEPVPOVETEKVEAQKAEVVLAKYTDSSLKXANATET 753
Db      760 MVDELIGOSSVTRTEITKVNASFSLDSVDYTKVAINVLSDAAV--NLNDANLKIYEKKT 818
Qy      754 LAGLRNNLTLQIMDNNSIWAEEKLLA 780
Db      819 -----NTLVKTV-----PLHGDFEKLMS 836

```

RESULT 41
 T38495
 hypothetical protein SPAC29B12.07 - fission yeast (Schizosaccharomyces pombe)
 C1Species: Schizosaccharomyces pombe
 C1Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
 C1Accession: T38495
 C1Genes: S.; Church, C.M.; Bartell, B.G.; Rajandream, M.A.; Wood, V.
 Submitted to the EMBL Data Library, September 1997
 A1Reference number: 221797
 A1Accession: T38495
 A1Status: preliminary; translated from GB/EMBL/DBJ
 A1Molecule type: DNA
 A1Residues: 1-1969 <GEN>
 A1Cross-references: EMBL:Z99164; PIDN:CAB16252.1; GSPDB:GN00066; SPDB:SPAC29B12.07

A1Experimental source: strain 972h-; cosmid c29B12
 C1Genetics:
 A1Gene: SPDB:SPAC29B12.07
 A1Map position: 1
 A1Introns: 664/2

Query Match 3.6%; Score 148.5; DB 2; Length 1969;
 Best Local Similarity 19.6%; Pred. No. 6.8;
 Matches 179; Conservative 113; Mismatches 288; Indels 331; Gaps 46;

```

Qy      14 ENNRVSYIDGQATQKTNLTPEDEYSKKEGINAEQIVIKITDQGVTSHGHHYHNGKV 73
Db      237 EVHEVPSIHGEGTETVNLTSAGNEKE--FNDSVVLNTT----- 276
Qy      74 PYDAIISSELMKQDNYLKDEDIYNEVKGVIVIKVDQKYVYLKDAHADNVRTKEIN 133
Db      277 FRENQISS-----LDVVDKLEDEKPVSN-----DNIKANITSS 308
Qy      134 ROKQESHQREBGTIPRNDGAVALARSQGRYTTDDGYIFNADLIEDTGA-----YIVPH- 188
Db      309 TTVDRASSQLSENNESFFPHNQNQSSQHQITLD---FVASKPVYDTPELSKENTLVSE 365
Qy      189 --GDHYHYIPKNELSASELAALAE--PLSGRNLNSRTYRQNSDNTS--RTNWVPSVN 243
Db      366 NLMDPKPMLPETEVDFGEPLVEKAPSFYV--QNSSAVQPTTKTSEDSDYDLHTNEIPYVAQ 424
Qy      244 PGTNTNTSNSTNSQASQNSDIDSLKQLYKLPLSQRHVESDGLVFPQAQITSTARG 303
Db      425 P-PSSFEKENNKDYSK--LENPDISS-----SPLS----- 451
Qy      304 VAVPHGDHYHYIPYSQMSLEERIRIIPLRYSNHWVPDSPEQSPQPTPEPSPGFP 363
Db      452 ---PTEDLFPRNDP-----EENLFSALGLNSN---TGQPSRSTQSKPSIDPESITV 498
Qy      364 APNLKIDS-----NSGLVSQLVKRVGEGVFEKEGISRYVF---AKDLPESTVKLE 412
Db      499 TDNQ--DILLPSQLTNALQAEANATKVSSENTINDELLIDSEFTLSMNFLESSYVQT-- 554
Qy      413 SKLSKQESVSHITLVA-----KKNV-----APRDEF-----YKAVNL-----L 447
Db      555 NKYLKSSASPPANAPVSSDVHKNENAGTARRAQAQSAFASKAKYSSPYDLPEIYQV 614
Qy      448 TEAHKALFXKNGRNDPQALDKLERLNDESTNKEKLVDDLLAFIAPITHEPRLKPSNQ 507
Db      615 AQQKRSVQSNNRQYSPFRP-----ATSPNPRSLPPEPBG 651
Qy      508 I---EYTEDVRIAQLADKYTSDGYIFDEHDIISDEGDAYVTPHMGH----- 552
Db      652 VNAPWSQTPNPISAEPIRRPATTVGH---TPNLVSPKNTYNSRHMAYEMTKSHINVIS 708
Qy      553 -----SHWIG-KDLSDEKRVAAQAYT-----KEKGILP--SPQAD 586
Db      709 PGPSLQVNAPYTPTSGELGNRVSNPTKEFVSTSSAAPANTRMALIREPGILSPUSP--- 765
Qy      587 VKANPT-----GDSAAIYNRVYKGEKRIPLVRLP---YVWEHTVE--VKNGNLI 631
Db      766 -RVQPLVSRRESIISMSSASSY-----VLELIPRPMSSLEHTNMSMSSGNL-- 813
Qy      632 PHKDHVHNKIFAMFDDHTYK--AEQGYTLEDLFATIKYVVEHPDERPHSNDGWNASEH 688
Db      814 ---QRTANL-----YKPMTPFNAYNIK----- 832
Qy      689 VLGGKDHSEDRNKPFADEBPVEETPAE--PEVPOVETEKVEAQKAEVVLAKYTDSSLKXANATET 740
Db      833 -----NSNQRETKYPYQPAIVSEYTOGSSSLPTSGEENAIINSPGFTPLAA 881
Qy      741 VTDSL--KANATETLAGLRNNLTLQIMDNNS-----IMAEKLL--ALLGS 785
Db      882 QKDATTIYPSHAQATLYG-----MNDNDRDNEGHIDILQSDMEVLPHPHNSAYHAN 933
Qy      786 NPSSVSKERIN 796
Db      934 APVSGHSEGLN 944

```

RESULT 42

hypothetical protein PF0765w - malaria parasite (Plasmodium falciparum)
 CSpecies: Plasmodium falciparum
 C.Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
 C.Accession: E71606
 R.Gardner, M.J.; Tetteh, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
 ; Perter, M.; Salberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
 Science 282, 1126-1132, 1998
 A.Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
 A.Reference number: A71600; PMID:9804551
 A.Accession: E71606
 A.Status: preliminary; nucleic acid sequence not shown; translation not shown
 A.Molecule type: DNA
 A.Residues: 1-980 <GAR>
 A.Cross-references: GB:AE001362; NID:G3845271; PIDN:AACT1942.1; PID:G384527
 A.Experimental source: clone 3D7
 C.Genetics:
 A.Gene: PF0765w
 C.Superfamily: hypothetical protein MJ1322

Query Match 3.6%; Score 148; DB 2; Length 980;
 Best Local Similarity 18.4%; Pred. No. 2.7;
 Matches 143; Conservative 131; Mismatches 303; Indels 200; Gaps 34;

```

OY 90 YKLK-----DEDIVNKGKGIYKDGKYYV-LKDAHADNVRKEEINROK--QEH5 140
DB 47 YELKEKLNDEKIID-----LKKNSQYVKMDYEREN-NLINEINKLPIENK 99
OY 141 QHREGTFRNDGAVALARSGRYYTDDGIYFNASDIIEDTGAVYIVPHGDHYIIPKEL 200
DB 100 MIVENGEMNNKLEEMQOKR-----ELINNLDISDELAKCT-----EQVNSISRN-- 147
OY 201 SASLEAAEAPLSGNGNSNSRTYRONSNTSRTNWPVSYPST-----TNTNTS 252
DB 148 ---NAVE---KEKENIINELQILRMKND-TMRKRISKFVQEKVLFKLYTLNNDIFS 199
OY 253 NNSNNSQASQSDNDISLKLQ-----YKPLSQRHVESDGLVDPAQITSRTARGAV 306
DB 200 KNEKLNQOKLNDVNEKYNIVECLNNK---TEHK-----QIEKKIERINTL 246
OY 307 PHGDHYHIFPS-QMSELEERIARIIPLRYGNHWVPSRPPQSPPTPSPSPQAP 365
DB 247 KONYIYLLKKEIDLKKELEKNIHGGKLEHLSHC----- 281
OY 366 NLKIDSNSLSQLVKVGEGYFEEKGISRYVFAKDLPSFTVKNLESKLSQESVSHL 425
DB 282 ---YEENQKLNBEIKR-----NSFIKKKRDILDTNIEMLKKKEINNINIK 326
OY 426 TAKKENVAPROQEPFD-KAYNLLTEAKALFXNK--GRNSPOLDKLELNDESTNK 481
DB 327 LMEKONVAKNNEQILKDIKDENEKNENHVNKLONELKELONKISKIDIECKEKEDK 386
OY 482 EK-LVDDLLAFAPITPRLGKPSQIEYTEDE-VRI-AOLADYTSQDGIYFDEHDI 538
DB 387 IKNLDDLL-----EKKKCIENLKDELINIKKKMEDM-----HMTNEMDL 428
OY 539 SDEGDAYVTPHGSHWIGKDSLDEKVAAYTKEKGIPLPSPADVYKANPTGDSAA 598
DB 429 SNK-----VEELNRIKNTYKKNIVELNDELDTIKKLNDE--- 463
OY 599 IYNNVKGKRIPLVLPYV-EHYVEVKNK---NLIIPKCHYNNIKFAMDHTYAP 653
DB 464 --EPLKEBKKNIMVYKIKYKEIQIKENEISLKKNEQNLHLVKNBEINKEIILK 521
OY 654 NGYTEDLPATIKYVVEHDERPHSNDGKNASBEVLGKKHSEDPNNK----- 702
DB 522 NKYDEIMNITIQYKKKIOBEKMDLNKIKSMDQTHKQIOEEMQENKELKRLKNVCDM 581
OY 703 -----EKADPEVEETPAE-----PEVPOVETEKEVAOLKEAEVL-----A 739

```

DB 582 NLQSQIILIKENKMOEVEEYKNLLKQKQDELKNIIOEYDERIEIONKEMEDIVNDCEE 641
 OY 740 KYTSSLSKANTTELLAGRNLLTLQIMDNNSIMAEAKELALLKGSNPSSEKIN 796
 DB 642 KLQAKINNKLLTATNNANNNN-LMDEN-LKEKDKINDLMKMEK--KKEEIN 693

RESULT 43

hypothetical protein F49E2.5a - Caenorhabditis elegans
 CSpecies: Caenorhabditis elegans
 C.Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C.Accession: T22458
 R.Sulston, J.
 submitted to the EMBL Data Library, October 1994
 A.Reference number: Z19566
 A.Accession: T22458
 A.Status: preliminary; translated from GB/EMBL/DBJ
 A.Molecule type: DNA
 A.Residues: 1-1230 <NLL>
 A.Cross-references: EMBL:Z46267; PIDN:CAA86428.1; GSPDB:GN00028; CESP:F49E2.5a
 A.Experimental source: clone F49E2
 C.Genetics:
 A.Gene: CESP:F49E2.5a
 A.Map position: X
 A.Introns: 37/2; 66/3; 210/3; 284/3; 367/3; 411/3; 552/3; 737/3; 767/3; 790/3; 853/3; 90

Query Match 3.6%; Score 148; DB 2; Length 1230;

Best Local Similarity 18.4%; Pred. No. 3.7;
 Matches 177; Conservative 123; Mismatches 351; Indels 312; Gaps 40;

```

OY 8 QARTYKENNRVSYDQKQATOKTENLTPDEVSKEGINAEOIVIKITDQGYVTSQHYH 67
DB 79 QKQIAKHAEQKYNNAKKAKEKARBAEAKKGAAGOEHEIKMK----- 123
OY 68 YNNGKVPYDALISELLMKDPNYKLD-EDIVNKGKGIYKDGKYYVILKDAHADNV 126
DB 124 -----ABQRIQKEQKKEADLKQAEKKKEKAVKAE-----KAEKAEKT 164
OY 127 R-----TKELNRQKQEHQHR-----EGTFRNDGAVALARSGRTTDDGIYFNASD 175
DB 165 KKAETPAVEBEIIVKQVANDRSAPAPPEKTPNTPAEP-----AEQ 207
OY 176 IIEPTGDAYIIVPHGDHYIIPK--ELASLEIAAAEAPL----- 212
DB 208 VQBITG-----KKNKKKKKSSSEKTAAPASVQVVEQKVTTEPHQQA 253
OY 213 -SGRGNLSNSRTYRONSNTSRTNWPVSYPSTNTNTSNNNSQASQSDNDISL 271
DB 254 PQEKKKKKKKSSSENVPASETPVEPVVE---TTPPASENQKKKK----- 298
OY 272 KQYKPLSQRHVESDGLVPPAQITS-RTARGAVAPRHGDHYHFI-----PYSQMSF-- 322
DB 299 -----DKKSSSEKVEEPVQAEAPSKKPTADDMDLDTVTAKEEKDPAETPA 350
OY 323 -LEERIARIPLRYSNHWV-----DSRPEQSPPTPSPSPQAPAPNL 367
DB 351 APVEEVENVENVENKSTTPPATENKKKKKKKSESEKVTQAPV-ESAPAPQ----- 405
OY 368 KIDSNSLSQLVKVGEGYFEEKGISRYVFAKD-LPSETVKNLESKL-----SKQESV 421
DB 406 -----VEGVVETTPPASENKKKN-----KDKKKSEBEKAVEEPVQAPASSKPTA 451
OY 422 SHTL-----TAK--KENVAPROQEFYDAYNLLTEAHKLFYNNKGRNSPOLDKLE 472
DB 452 DDSMDFLPVDPAKDRBEVAPVAVAKVDETAATSENKKKKDKKKSESEKAVE--E 508
OY 473 RLNDSTNKEGLVDDLLAFAPIT---HPRLGKPSQIEYTEDEVRIAOLAQKYTT-- 526
DB 509 PVQAAPISSKKTADSDMDFLPVTAKERVBEVAPVQVQVEQKNSSESSEPVAKLITVS 568
OY 527 -----SDGIYFDEHDIISDEGDAYVTPHGSHWIGKDSLDEKVAAYTKEKGIPLPS 582

```

Db 569 NTEASAVNNMGESDITVPADEVIT-----ODPVSAKQEVLPHEVPS--IPEE 615
 QY 583 PDADVANKPTGDS-----AAAIYNRVKGEKRIPLVRLP-----YMVHTVEVKN 626
 Db 616 PVA-VSKRPTADSMDFLDPVTPTEASTSEAPAVSKPTESIEDLEIVTYHADV-T 673
 QY 627 GNLIIF-----HKDHYNIKFAMFDHTYAPNGTYLE--DLPA 663
 Db 674 GNTLSPSQSTPSNSVLLNGPQSKSKRHHKKKRTDSMSQSPSKEDLEFLFLH 733
 QY 664 TKYVEHF-----DERPHNDGWC-----NASEHYL----- 690
 Db 734 SEPKVKEKVPVAPVSKRPTADNDMDFLDPYTAKEKTESVEHELEAPMIVEPVHAENETA 793
 QY 691 ---GKKDHSDEPNKKNKKADEEPEVEETPAPEVEPQVETEKVQAOLKEAEVLLA----- 739
 Db 794 AAGGKKKKKK--NKKKKSS--ESTPAPPEVKEVPEIVEEVEFEKTTTPSTAAATAA 849
 QY 740 -----KVTDSIKANATETTLAQRNNLTQIMDNNSIMAEKLLALLKGNPSSV 790
 Db 850 PGSKKKKKKKKSSSSAPAEPEVKEVTEPEIVEEVEFEKTTVPSTAAAA-----PASA 903
 QY 791 SKE 793
 Db 904 SKK 906

RESULT 44

C95008
 Immunoglobulin A1 proteinase (imported) - Streptococcus pneumoniae (strain TIGR4)
 C:Species: Streptococcus pneumoniae
 C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #ext_change 24-Aug-2001
 C:Accession: C95008
 R:Retelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
 on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holteapple,
 neon, T.; Hickey, E.K.; Holt, I.E.
 Science 293, 498-506, 2001
 A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
 A:Title: Complete genome sequence of a virulent isolate of Streptococcus pneumoniae.
 A:Reference number: A95000; PMID:21357209; PMID:11463916
 A:Accession: C95008
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1856 <KUR>
 A:Cross-references: GB:AE005672; PIDN:AAK74260.1; PID:g14971537; GSPDB:GN00164; TIGR:SP4
 A:Experimental source: strain TIGR4
 C:Genetics:
 A:Gene: SP0071
 C:Superfamily: Streptococcus sanguis Iga-specific metalloendopeptidase

Query Match 3.6%; Score 148; DB 2; Length 1856;
 Best Local Similarity 19.8%; Pred. No. 6.7; 287; Indels 242; Gaps 38;
 Matches 158; Conservative 109; Mismatches 242; Gaps 38;
 QY 107 IKVDGKYVY-YLEDAAHADNVRTKEINROKQESQHREGGTPRND-----G 152
 Db 172 LKEIGVYIGYIK-TKKQNTLSRTVDKYSQNRQSNSTSTSVHASDLEMNQGQ 230
 QY 153 AVALASQGRYTTDGYIINASDIIDTDGAYIVPHGDHYHYPKNELASGLAAAEPL 212
 Db 231 KYSL--QGEASGDDG-----LSEKSIADNLSNDSPA 262
 QY 213 SGRGNLSNRTYRQNSDNTSRNWPVSINPCTTNTNNSNTMSQASQSDIDSLK 272
 Db 263 S-----QVEQPHDHGSESVRPTVPEQGNPVSAITVQSAEEVLATTND-----R 307
 QY 273 QLYKPLSORHVESDLVFD-----PAQITSRITAGVAVPHGDHYHFLPYSSOMSELE 325
 Db 308 PEYKPLFLETKGTQEPHGEBAAVREDLPVYTKLETKGTQOGPHF-----GEAVRFE 360
 QY 326 RIARIIPLR-----YRSNHWPDSRPQSP--QPTPEPSPQPA-----PNLKI 369
 Db 361 EAYTEPLATKGTQEPGHGKATVRETELEYTEPVATKGTQEPHGEBAVEEELPALE 420

QY 370 DENSSLVSQLV-----RKVG-----EGYVEEKGISR 396
 Db 421 TIRNRTEIINIIYTTTEIQTPTLLKNRRKRIEBOGAGRTTIOYEDYIVANGVETKEVR 480
 QY 397 YVFAKDLPEYTKNLESKLSKQESVSHLTAKENAVAPRQEPYKAVLL--TEAHKAL 454
 Db 481 TEVA--PNEVAVKGTLVKVPVETITLTKVEN-----KKSITVSYVLIPTTSAYVS- 531
 QY 455 FANKRNSDFQALDKLERTLNDESTNKELVDDLAFLAPI--TH-PERLKPNSQIY 510
 Db 532 ---AKQVPHG-DKLVEVDIENPAKQVIGS-LDYTPYTVKTHLYNLLENNE--EN 583
 QY 511 TEDEVRIQDLADKYTTSDGIFDEHDIISDECDAYVTPHMGSHMIGKSLDKEKVAQ 570
 Db 584 TETSTQDPQL--EYKIEIKDIDSVELVQKENDRI-----RRLL--SLSEAPIDYK 631
 QY 571 AYTKERGLPPSPDADVANKPTGDSAAAIYNRVKGEKRIPLVRLPYMEHT-----VEVK 625
 Db 632 YFVKYK-----SDRFKEMYLPVKS-----ITENTDGTYKVVA 664
 QY 626 NGNLIIPHKHYNIKFAMFDHTY-----KANGYT-LEDLPATIK-----YVEHP 672
 Db 665 VDQVVEEGTDGYK-----DDYFTVAKSKAQEGVYTSFKQLYTAMQSNLSGVYTLAS 717
 QY 673 DERPHNDGWNVASEHVLGKQDHS-----EDPNKPKADE--EPVEETPAEPE 718
 Db 718 DM-----TADVSLQDKQTSYLTGAFTGSLIGSDGTSYALYDLKPLPDTLNGAT 768
 QY 719 VPQVETEKVQAOLKEAEVLLAKVYDSSL-----KANATETTLAQRNNLTQIMDNNS 770
 Db 769 VADLDIKTVASDSKENVALAKAANSANINNVAVEGKISGAKSVAGLVASATNTYIENS 828
 QY 771 IMAEAKLLALLKGSN 786
 Db 829 FTG---KLIANNQDSN 841

RESULT 45

A56976
 transfer complex protein Trst - Staphylococcus aureus
 C:Species: Staphylococcus aureus
 C:Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #ext_change 15-Oct-1999
 C:Accession: A56976
 R:Morton, T.M.; Eaton, D.M.; Johnston, J.L.; Archer, G.L.
 J. Bacteriol. 175, 4436-4447, 1993
 A:Title: DNA sequence and units of transcription of the conjugative transfer gene comple
 A:Reference number: A36891; PMID:93322322; PMID:7687249
 A:Accession: A56976
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-700 <MOR>
 A:Cross-references: GB:L11998; NID:9310606; PIDN:AAA71960.1; PID:9310616
 C:Genetics:
 A:Gene: trst
 C:Superfamily: DNA topoisomerase I topa

Query Match 3.5%; Score 147; DB 2; Length 700;
 Best Local Similarity 20.5%; Pred. No. 1.8;
 Matches 101; Conservative 84; Mismatches 174; Indels 134; Gaps 22;
 QY 388 VFEKGISRYVFAKDLPESTVKNLESKLSKQESVSHLTAKENAVAPRQEPYDKAY-NL 446
 Db 96 IFKENKIDEVITATD-PAREGENIAYKILNQAKVDKTYIKKLMVLSKVSSIRKAFKNI 154
 QY 447 LTEAHKALFXNKGNSDFQ-----ALDKLERLNDESTNKELVDDLAFLAPIHP-- 498
 Db 155 LPKEKTYGFKGGRARELSMDWLVGINLSRHFKIRF-----LQNDGYTHIGRVSSPTL 208
 QY 499 -----ERLGENSQIYEYDEVRINQDLADKYTSP--GYTDEHDIIS- 539
 Db 209 NNVYRNENNIKQFKGKKFVKSATINKDEOEVK--TELKPKFSDSEDLHEFLF-ENDITDL 266

```

QY 540 -----DEGDVAVTPEMHGSHWIGKDSLSDKEKVA-----AAAYTEKGIPLPSP 583
D 267 TOKGLVTIDIEKEIGVTMPKPYDLSALOEDNMNDYKISAKRTLEIAQLUYEKLLTYRT 326
QY 584 DADVANKPTGSAALAYRVKGEKRIPLVRLPYMVEHTV-----EVKGNLIIPKX-DH 636
D 327 DSRV-----ITDEKEMLEENIDYLEKEITKINLNNELTNLSLNPFSKIDH 372
QY 637 Y-----HNIFKPMFDHTYKAPNGYTTLEDJFATIKYYVE 670
D 373 YAILITGMDPNKVDLKESEINVTYKSLQNVAMNPNDKQYETT---TTEIAVKKLMFEVK 429
QY 671 HPDERPHSNDGNGNASEHVLGKKDSEDPNKNPKADEE--PVEETPAEPEV--POVETEK- 726
D 430 GKIIQ-----DNGFKALLNKQKTSEETIPNFEKNEEVDIEDLLEKETTPPKRYTEKT 482
QY 727 -VEAQLKEAEVLLAKVTDSLK-----ANATETTLAQLRNNLTLOIMDN-----NSI 771
D 483 LKKMANPIETLEBGLKSTLEKVGKLGTPATRADIIENLKKNKYIQVQKNKIYTKNGI 542
QY 772 MAEAEKLLALLKG 784
D 543 LA-----CLLEG 550

```

RESULT 46

F84730

probable myosin heavy chain (imported) - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cress)

C/Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C/Accession: F84730

R/Lin: X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanKen, S.E.; Umayam, L.; Tallon, L.

euse, D.; Niernan, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.

Nature 402, 761-768, 1999

A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A/Reference number: A84420; MUID:20083487; PMID:10617197

A/Accession: F84730

A/Status: Preliminary

A/Molecule type: DNA

A/Residues: 1-1269 <STO>

A/Cross-references: GB:A8002093; NID:G6598483; PIDN:AAC69932.2; GSPDB:GN00139

C/Genetics:

A/Gene: At2g32240

A/Map position: 2

Query Match

3.5%; Score 147; DB 2; Length 1269;

Best Local Similarity 18.4%; Pred. No. 4.4;

Matches 160; Conservative 121; Mismatches 293; Indels 294; Gaps 34;

```

QY 70 NGKVPYDAIIBELLMKPNYKLNQEDIVNEVKGKGYVKKDQKYYVYLKDAHADNV--- 126
D 33 NGEVP-----KEE---KEEDEEFLKEVEKADAKDADKADHVEVE 71
QY 127 RTKEINR-----QKQEHSGHREGTTPRNDGAVALARSGRYTTDDGYIFNASDI 176
D 72 EOKVEIHESSSGSQRELHESQEKALELELERVAGELKRVSEHTHLKDE--LLSAEX 129
QY 177 IEDT---GDAYIVPHGDHYHYPKNELSASELAAAEFLSGRGLNSRTYRRONS- 230
D 130 LEETEEKKGDLEVVQKQOEKIVEGEERHSQKSLDALOSHDAKDELTEVKEAFAL 189
QY 231 ---NTSTNVVPSNSGCTNTNTNSNSNSQSGSNDIDSLKQYKPLSGRHES 286
D 190 GIELESSSKLLI-----ELEEGKRSSEKQKEELKQ-----SASHADS 230
QY 287 D---GLVF-DPAQITSRFAGVAVPHGDHYHYPYQSELESEARIITPLRYSNHWP 342
D 231 ESQALRESELLKSTKESAK-----EMEKKASL-----QDEIKLNL 267
QY 343 DSRPEQSPQPTPEBSPGPQPA-----PNLKIDSNSSLVSQLRVKGEYV 388
D 268 EKMSENEKVAALKSSAGELAAVQELALSKSRLLTETQKVSSTFALIDELTQELQCKA 327

```

RESULT 47

T21559

hypothetical protein F29G6.3a - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000

C/Accession: T21559

R/Harris, B. submitted to the EMBL Data Library, August 1996

A/Reference number: Z19441

A/Accession: T21559

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1929 <WIL>

A/Cross-references: EMBL:Z78543; PIDN:CA801755.1; GSPDB:GN00028; CESP:F29G6.3a

A/Experimental source: clone F29G6

C/Genetics:

A/Gene: CESP:F29G6.3a

A/Map position: X

A/Introns: 10/1; 43/3; 61/3; 88/3; 110/3; 144/2; 183/3; 218/3; 252/3; 288/3; 1083/3; 1210/3

Query Match

3.5%; Score 147; DB 2; Length 1929;

Best Local Similarity 17.6%; Pred. No. 8;

Matches 150; Conservative 125; Mismatches 331; Indels 248; Gaps 42;

```

QY 12 VKENRRSYIDGKATQKTENLTPDEV-----SKREGINAEQVIYIKITTDQYVYSHDHY 66
D 661 IAQSPQVQTIIDYFKEINRAHHPVSTPPSLFNNRRESFHHHSFPLNSLEERKRSYVERA 720
QY 67 HYVNGKVPYDAIIBELLMKPNYKLNQEDIVNEVKGKGYVKKDQKYYVYLK----- 118
D 721 HHY-----LBDSPANVVDQKRIIE-----HHYVPLEKRSQSLD 754
QY 119 -DAAHADN-----VRTKEINRQKQEHSGHREGTTPRNDGAVALARSGRYTTDDGYIFN 172
D 755 LKVVHPHHYVPLVRKTTTEHDQKVNHHHFIPEKTEA----- 794

```

QY 173 ASDIIEDGDAYIVDH---GDH-----YHYIP-KNE--LSASEL-----AAA 208
Db 795 VKDVVHNNHYIPLAHKTKDEHNVNITREHHIYIPLKTEVOQSKELEERINQVEMPTTKHET 854
QY 209 EAPLSGR-----GNLSNRT--YRRON-----SDNTSRTNWPSVSNPGT 246
Db 855 KSPFTGRSHSYPIPGVNEQSKTEIFRQHNFPVVDKKEKVAEELINQGHNVVPAHSNVAV 914
QY 247 TNYTTSNN-----SNTNSQAGSNDIDSLKOLYULPLRSQHNVEDGLVFPRA---- 294
Db 915 SDKERPISEHHNFVPLATKTVETBETHSND-----LSHHNMPPLPAREBEKTAAGKVGGEH 970
QY 295 ---QITSARGVAVPHGDHNYFIPIYSQMSLEERIALIPLRYSNNHWDSRPEQSP 351
Db 971 HYPLASTER-VSYNHQGEHHYVPL--IAKPTSHMSKIVTIQH--YIPLANRHOQVP 1024
QY 352 QPTPEPSGP---OPAPNLKID-SNSSLVSOVLVRKVGEGVEEKGISRYVPAKDLP--- 404
Db 1025 RLPSHHNYIPLAINDHDSHQKVDLTNHYQLPPIVNA-----EQKEAKVLEHQYVPLRP 1077
QY 405 -SEYVKNLESLKSQESVSHTLTAKKENAVAPRDOEFYDKAVNLLTEAHKALFNKGRNSD 463
Db 1078 KKDSEFANYSKAPQLPREHHYYPAPVRSVEHOKKHVHQ--HNYIIPAVH----HHETNHQ 1131
QY 464 FQALDKLERLNDESTNKEVLVDLLAFIATITHERERKGRNSQLEYDEVRITQALADK 523
Db 1132 LDRHGYPVPAKHEAKKNVUYEH--QYIPRTTEQTNQYDRRNY---VPVGKTA-K 1184
QY 524 YTSDDGYIFDEHDIISDEGDAYVTPRMGSHWIGKDSLDEKKAQAQAYTEKGLPLPSP 583
Db 1185 ATTTDINHNOHQVPOASPRKTISEIRH-----KTKQNVVNEHGVPEPVK 1231
QY 584 DADYKANFTGDSAAIYNRVKGEKRIPLVRLPYVEHT-----VEVKNGLIIPKHQHY 637
Db 1232 KSDKTSYV--DHSKHXY-----LPSVY-RTESKODFGKSHG--YVPSLEHS 1273
QY 638 HNIKFAWDEHDTTYPAR-----NGY-----TLEDLFAITIKYY 668
Db 1274 QRIRGVVSTNNVYPPVTIKIEKNASDPTKAGYLPQSRKSKSDQIDHSHKGFPLPSIKHQ 1333
QY 669 --VEHPDRPHS-----NDGKNASEHYLKGKHSDEPNKRFKADSEPVVE-TPAE 716
Db 1334 EAQHDADSHRGYVPTPKKANNDYWGFERDHHYVPDTLKKDKIKQIHRHNFPVSVPAPK 1393
QY 717 PEVPQVETEKVEAQ 730
Db 1394 PD-----EKFEQG 1401

RESULT 48
F95133
immunoglobulin A1 protease [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 14-Sep-2001
C:Accession: F95133
R:Teiltein, H., Neilson, K.E., Paulsen, I.T., Eisen, J.A., Read, T.D., Peterson, S., Helton, J.D., Umayam, L.A., White, O., Salzberg, S.L., Lewis, M.R., Radune, D., Holtapple, T., Hickay, E.K., Holt, I.E.
S:Science 293, 498-506, 2001
A:Authors: Loftus, B.J., Yang, F., Smith, H.O., Venter, J.C., Dougherty, B.A., Morrison, A.: Complete Genome Sequence of a Virulent Isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: F95133
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-2004 <KTR>
A:Cross-references: GB:AA005672; PIDN:AAK75263.1; PID:G14972632; GSPDB:GN00164; TIGR:SP2001
A:Experimental source: strain TIGR4
C:Genetics:
C:Superfamily: Streptococcus sanguis Iga-specific metalloendopeptidase

Query Match	3.5% Score 146.5; DB 2; Length 2004;
Best Local Similarity	19.7% Pred. No. 9;
Matches 173; Conservative 116; Mismatches 306; Indels 281; Gaps 48;	
QY 34 TPDEVSKREG-INAEOIVIKITDGVY-----TSHGDHYHYNGKV--PYDAIISELLM 85	
Db 489 TPVEKTEKQGEPEKTEEVVVKPTEETPVNPNEGTTEGTSIGEAENPVQPAEESTTNSEKVS 548	
QY 86 KDPNRYKLKDEIIVNEVGVGVIKVDGKYVYIYIKLAALADNVRTTEELINROKQESHQREG 145	
Db 549 PDTSSKKTGEVSSNP-----SDSTTSVGESENKPEHNDSEKNENSEKTEVE 591	
QY 146 GTP--RNDGAVALARSO-----GRYTTDDGYIFNADSIIEDTDAYIVPHGDHYHY 194	
Db 592 EVFVNPNEGIVEGTSNOETEKPVQPAEETQTSKIAN-----ENTEVSNKPSDSK-- 643	
QY 195 IPKNELSASELAIAEAFLSGRGLSNSRTRYRQNSDNTSRIN--WVSVSNGP-----TT 247	
Db 644 PPEVSENGPE-----KNGTARKPKNSGNTTSENGQTEDEPNSGNTEDVST 689	
QY 248 NNTNSNNNSNNSQASQSDNIDSLK-----QLYKP--LSQRHVESDGL 289	
Db 690 ESNTS-NSNGNEELKQENELDPDKVPEPEKTELELRNVSDLEIYSLNGTYKQHSILEQV 748	
QY 290 VFDP-----AQTSRTAGVAVPHODDHNFIPIYSQMSLE-----EELIATIP 332	
Db 749 PSNNSYFVKKVSSSPFDV-----YLPVASISEERKNDKILYITAKVKEGLQOEIE 799	
QY 333 LRYASN--HWVPDRPQPSPPQTPPEBSPGPQAPFNLIKIDNSLSVLQVRYKVEGYVE 390	
Db 800 SRVNDNTFYIAKKGTGE-----TTNFTSFENLKAI-----N 832	
QY 391 EKGISRYVPAKDPSEFYVKNLESLSKQESVSHTLTKKENAVARDOEYDKAYNLLTEA 450	
Db 833 QNPSTGYTHLAASLANEV--ELCPDRERSYIKQFTGRL--IGEKDKNY-AIYNL----- 882	
QY 451 HKALFNNKGG-----RNSDPALDKLERLNDESTN--KEKIVDDLAFLAFLPITHR 498	
Db 883 KKPLFENLSGATVEKLSLKNVAISGKDIDSLANEAOONTKIKQVHVDGLA-----G 935	
QY 499 ER-----LGKNSQIEYTEDEVRIAOLADKYTTSDDGYIFDEHDIISDEGAYVTPHNGH- 552	
Db 936 ERGIGGLAAR-EQSSITSESK-GRINIYETTAAT-----NIGGM 975	
QY 553 -SHWIGKDSLSDEK-----VAAQATYEEK-GILPSPDDADVKANPT--GD----- 594	
Db 976 VGHLTGPKALLTKSKATVAIISNTNTSDQYVGLAGLVDRDAQIQDSVABEDINNVKHFQ 1035	
QY 595 ----SAAAIYRVYGEKRIPLVRLPVWEHT-----VEVKNQULLIPHKDHYNIK 641	
Db 1036 RVAGVAGNLDRTSGD-----VRHAGSLITNVLSDNVNTNGAALTCY--HYNEMK 1082	
QY 642 FAWPDDHTYKAPNGYTLLEDLFATIKYVVEHPDERPHNDGKNASEHYLGKDKHSEDPNK 701	
Db 1083 VK--DTSSSKANRYNVN-----TL-----VKDEVSKESPEE--R 1113	
QY 702 NFKADEEPVEETPAPE--VPQVETEKVEAQLKEAEVLLAKYTDSSLIKANAT-----ETL 754	
Db 1114 GTMLDAQIASKKAEINPLIPTPELPISTSGK-KDSD--PSKAVYQAQKNLLTYKNIETKL 1170	
QY 755 AGLRNNLTLL--QIMDNNSIMAEAKLLALLKGSN 786	
Db 1171 LPFYNAKATIVKYGNLVNENSLLYOKELLSAVMMKDN 1206	

RESULT 49

S67087

hypothetical protein YOR195w - yeast (*Saccharomyces cerevisiae*)

N:Alternate names: hypothetical protein O4806

C:Species: *Saccharomyces cerevisiae*

C:Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 19-Apr-2002

C:Accession: S67087

R:Hughes, B.; Pohl, T.M.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S66685

A:Accession: S67087

A:Molecule type: DNA

A:Residues: 1-821 <HUG>

A:Cross-references: EMBL:Z75103; NID:g1420464; PID:g252389; PID:g1420465; GSPDB:GN00015;

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:SLK19; MIPS:YOR195W

A:Cross-references: SGD:S0005721

A:Map position: 15R

Query Match 3.5%; Score 146; DB 2; Length 821;
Best Local Similarity 19.2%; Pred. No. 2.6;
Matches 136; Conservative 126; Mismatches 269; Indels 178; Gaps 32;

```
QY 225 RRONSDNTRTWVSVPNGPTTNT--NTSNNSTNSQ-----ASQSDIDSL 270
DB 19 REONSSENSOER-----NPRTFNSEPDSPSSFGSVIHHPHLEKEKEDKODLDRS 72
QY 271 LKQVTKLPLOSRRH-----VESDGL--VFPDAQITSTR--ARGVAVPHGDHYHPIYS----- 318
DB 73 IDYGSLSALNNKNNANPLENIDINKMFDDKSDSGINDKGGASTSDGVALLANYSPIRV 132
QY 319 QMSLEERIIARIIPLYRSNHNWPDSPRQPSQPPTPEPSGPQAPNLIKIDSNLSLSQ 378
DB 133 EMNSEK-----KSDKNVDVDENDKESGSHNKKLKLQLESVPDLKOSTDIIND 182
QY 379 LVKRYGEGVFE--EKGISRYVF-----AKDLPSEYVKNLESKLSQESVSHLT 426
DB 183 KEELINSPMAIDMETNISPNKFIINDGYERNDSFNINTDLK--LENNINEKQREDEIK 241
QY 427 AKKENVAPRDEFFYVKA-----YNLL-----TEAHKALFNKGRN----- 461
DB 242 SNNNNVAINIDNAYKKEDEBENDITNSHINRLTPLYETSARESNSVEGENDYDDNQJDI 301
QY 462 --SDFQALDKLERLNDESTNKEKLVDDLK-----FLAPITPERLG 502
DB 302 RHDNQIVAKNNEBLTDQIYHLNQMLNLSISKNESLSFOYEKLNKHQLLIDLTN--EKLD 360
QY 503 KPNISOIETDEVRJAQADKY-----TSDGYIFPEHDIIISDEGAYVTP--HMGHS 553
DB 361 KLN-----TERESDIAYK--EKFKKRIKEINTEIKVLNNOQKILQEPFASITEVNHIGE 414
QY 554 HWIGKDSLSDEKVAQAAYTEKGILPSPDADYVANTGDSAAIIVNRVGEKRIPIVR 613
DB 415 HENVTNLTQONEKI-----LNDKNVELENNKALKGN--NDKISEYETTLINDLSRIVQ 466
QY 614 LPYVVEHTVVEVKNGLIIPHKDH--YHNIKFAWFDDHTYKAPNGYTLLEDFATIKYVEHP 672
DB 467 LNDKIEST-----DIVLSKKNELDNKLSL--KETISISKDPNDSLLIQINELISTK 518
QY 673 DERPSINDGWNASHVILG-----KKOHSDEPNKFADEPVEETPAEPEV 719
DB 519 NNLQOKMDLNNLNDNKLKVQDKLIKNEETLKLKEAEIDSLNSEMDELKQKITSKDEF 578
QY 720 ----POVETKEVQOLKEAEV--LLAKYTD--SLKANAETELAGRNLL----- 761
DB 579 KMMOSKIEYTVDEAKIRNAEVTLENGDIEDLKESKLIHEITTELLENVYHKLNECELEK 638
QY 762 ----TLQI-----MDNNSIMAEKTLA-----LLKSNPSSVSEKRI 795
DB 639 QKFEKTSLELESLOKNSNIOAEHKELENHENLISLQNELKISSDRI 687
```

RESULT 50

S14032 kinesin-related protein cut7 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 02-Feb-2001

C:Accession: S14032

R:Hagan, I.; Yanagida, M.

Nature 347, 563-566, 1990

A:Title: Novel potential mitotic motor protein encoded by the fission yeast cut7(+) gene

A:Reference number: S14032; MUID:91015362; PMID:2145514

A:Accession: S14032

A:Molecule type: DNA

A:Residues: 1-1073 <HAG>

A:Cross-references: EMBL:X57513

A:Note: the authors translated the codon GCC for residue 2 as Gly and AGT for residue 10

C:Genetics:

A:Gene: cut7

A:Introns: 50/1

C:Function: A:Description: probably involved in mitotic spindle body duplication

C:Superfamily: unassigned kinesin-related proteins; kinesin motor domain homology

C:Keywords: ATP; microtubule binding; mitosis; nucleotide binding; P-loop

F:61-415/Domain: kinesin motor domain homology <KNOT>

F:147-154/Region: nucleotide-binding motif A (P-loop)

F:153/Binding site: ATP (Lys) #status predicted

Query Match 3.5%; Score 145.5; DB 2; Length 1073;
Best Local Similarity 18.6%; Pred. No. 4.1;
Matches 182; Conservative 136; Mismatches 357; Indels 301; Gaps 41;

```
QY 23 GKQATQTEMLTPPEVSKREGINAE-----QIVIKI--TDQGYVSHGDHYHYN 70
DB 147 GQTGTGTVMSCD--LSDSDGILSEAGALIPRALYQFSSLDNSNOEYAVK--CSYEELYN 204
QY 71 GKVYVDIISSELLMKDQPNYKLED-----INVEVGGVYIKYDKGYVYLKQAAHADN 125
DB 205 EEI--RDLVSEEL--RKPARVEFTSRGNAVITIGIEESY--IKNAGGRLRLRGSHRQ 260
QY 126 VRTKEEINROKQHS-----QHREGTPRNDGAVALARSGRYTTDDGYIFNASDIEDT 180
DB 261 VAATKCNDSLRSRHSITITLHRKVSSGMDETNSLTINN-----NSDDLK-- 307
QY 181 GDATVPHGDHYHYTPKNEISASGLAALAEAFLSRGMLNSRT----- 223
DB 308 ----ASKLMVDLAGEENT--GRSGAENKRRARETGMINOSLTLGRVI 349
QY 224 ----YR-----RONSNTSRNTWVSPVSNPGTNTNTSNNSTNSQASOS 264
DB 350 NALVEKAHHIPYRSKTLRLLODSLGGTKTSMTIVTS--STNTNLEETITSTLEVARA 406
QY 265 NDIDSLKQVTKLPLOSRRHVESDGLVFPDAQI-----TSRTARGVAVPHGDHYHPIYSQ 319
DB 407 K---SINKPQNNQVLVRKVLKDLVDIERLKNDLNAIRKNGVYLAESTYKSLMRVQ 463
QY 320 MSEI-----EEIRIARIIPLYRSNHNWPDSPRQPSQPPTPEPSGPQAPNLIKI----- 369
DB 464 NKDILCOEQAARKLEVLDLNVS-----SREQLQVSKSNQEHKKEVEALQQLVNSSTE 517
QY 370 ----DSSSLVSOQLVRKVGCVFEEKGISRYVFAKDLPEYVKNLESKLSQESVSH 424
DB 518 LESVKSNEKLNKLVLEIEGRKKYETNEAKITTVATDLSQYRESKEYIASLYEKDRT 577
QY 425 LTAKENVAPRDEFFYVKAAYNL-----TEAH 451
DB 578 ERNNKEN-----ENNFMNLKFNLLMLRSFHGSFTDETNGYFTLLNDFNASEBELNHSN 633
QY 452 KALFYXNGRNSDFOADKTLER-----LNDESTNKEKLVDDLKFLAPIT 496
DB 634 QLLISMRTKITEHFOSLDEALQASARSCAVPNSLILVSELOKSKNLLDLALHSLODIS 693
QY 497 --HPERLQKPS-----QIEYFEDE-----VR--IAQL 520
DB 694 MSSQKLGNGISSELLIELQDKMKESYROLVQELRSLYNLQHTHESSQKELMAYGVNDIAL 753
QY 521 ADKYYTS--DGYIFPEHDIIISDEGDAYVTPHMGSHWIKDSLSDEKVAQAAYTKKEGI 578
DB 754 VKTCTTSLNADIT--LSDYISDQSKRFESQODIANIGIKVSNFLQEQNESLYTK----- 808
QY 579 LPSPDADVKNPTGDSAAAI--YNRVGEKRIPIVR-----LPYVVEHTVVEVK 625
DB 809 ----ADILSHINDTNSNIRKANEIMNRSSEFLRNAAEQAEIVGANKERIOKTVE-- 860
```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 07:52:04 ; Search time 24 Seconds
(without alignments)
1559.719 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 4165
Sequence: 1 SYELGIYQARTVKNRNVSY.....KLIALIKGSPSSVSKKIN 796

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	171	4.1	1233	1 YFIE_YEAST	P43397 saccharomyc
2	161	3.9	1420	1 APX_XENLA	Q01613 xenopus lae
3	158	3.8	1118	1 UBP8_HUMAN	P40818 homo sapien
4	157	3.8	910	1 DNM MYCPN	P75354 mycoplasma
5	157	3.8	1658	1 YME7_YEAST	Q03661 saccharomyc
6	157	3.8	2195	1 SC16_YEAST	P48415 saccharomyc
7	156.5	3.8	1185	1 MAPX_DROME	P23226 drosophila
8	154.5	3.7	3924	1 ANK2_HUMAN	Q01484 homo sapien
9	154	3.7	1453	1 Y373_BOVIN	Q91423 bos taurus
10	153.5	3.7	1539	1 Y373_HUMAN	O15078 homo sapien
11	150.5	3.6	1176	1 YOH8_YEAST	Q08236 saccharomyc
12	149.5	3.6	1175	1 PTNL_RAT	Q62728 rattus norv
13	149	3.6	2869	1 RBPI_PLAIVB	Q00798 plasmodium
14	145.5	3.5	1085	1 CUT7_SCHPO	P24339 schizosacch
15	145.5	3.5	1240	1 YNJ1_YEAST	P53935 saccharomyc
16	144	3.5	1337	1 J1P1_MOUSE	Q96899 m c-jun-ant
17	143.5	3.4	900	1 SYA_MYCGE	P47534 mycoplasma
18	142.5	3.4	875	1 Z1P1_YEAST	P31111 saccharomyc
19	142.5	3.4	1066	1 STS5_SCHPO	Q74434 schizosacch
20	142	3.4	850	1 D7_DICD1	P54682 dictyostell
21	142	3.4	1744	1 TANA_XENLA	Q01550 xenopus lae
22	141.5	3.4	490	1 MYS1_YEAST	P54785 saccharomyc
23	141.5	3.4	1928	1 MYST_YEAST	P08964 saccharomyc
24	141.5	3.4	2314	1 PTP2_HUMAN	P23471 homo sapien
25	141.5	3.4	2748	1 NDM1_YEAST	Q00402 saccharomyc
26	141	3.4	914	1 IF44_YEAST	P39936 saccharomyc
27	141	3.4	1332	1 SPT7_YEAST	P35177 saccharomyc
28	140.5	3.4	516	1 P54_ENTFC	P13692 enterococcu
29	139.5	3.3	1609	1 LMGT_HUMAN	P11047 homo sapien
30	139	3.3	1957	1 SPOF_SCHPO	Q10411 schizosacch
31	138.5	3.3	1703	1 SNF2_YEAST	P22082 saccharomyc
32	138	3.3	1639	1 MSP1_PLAIVB	P04933 plasmodium
33	137.5	3.3	1790	1 USO1_YEAST	P25386 saccharomyc

34	137.5	3.3	2175	1 HMCU_DROME	P10180 drosophila
35	137.5	3.3	2459	1 MAPB_RAT	P15205 rattus norv
36	137	3.3	1176	1 PTNL_MOUSE	Q62136 mus musculu
37	136.5	3.3	679	1 YIS3_YEAST	P40563 saccharomyc
38	136.5	3.3	1183	1 CNA_STRAU	O53554 staphylococ
39	136	3.3	1630	1 MSP1_PLAIVB	P04932 plasmodium
40	136	3.3	1636	1 BUD3_YEAST	P25558 saccharomyc
41	135.5	3.3	1053	1 SLPM_BACBR	P06546 bacillus br
42	135.5	3.3	1976	1 MYHA_BOVIN	Q27991 bos taurus
43	135.5	3.3	3358	1 PCGV_MOUSE	O62059 mus musculu
44	135	3.2	1532	1 TGA_NEICO	P09790 neisseria g
45	134.5	3.2	960	1 YME6_YEAST	Q04279 saccharomyc
46	134.5	3.2	1124	1 TCP8_HUMAN	P37127 homo sapien
47	134	3.2	1163	1 RTN4_RAT	Q91811 rattus norv
48	134	3.2	1188	1 OSH1_YEAST	P35845 saccharomyc
49	134	3.2	6632	1 UN89_CABEL	O01761 caenorhabdi
50	133.5	3.2	1607	1 LMGI_MOUSE	P02468 mus musculu
51	133.5	3.2	1664	1 INT1_CANAL	P53705 candida alb
52	133	3.2	1189	1 YH6_YEAST	P47035 saccharomyc
53	132.5	3.2	1875	1 MLP1_YEAST	Q02455 saccharomyc
54	132	3.2	1164	1 TSCL_HUMAN	Q92574 homo sapien
55	132	3.2	2431	1 POLN_SRV	P08411 semliki for
56	131.5	3.2	764	1 MBS1_YEAST	P54199 saccharomyc
57	131.5	3.2	1046	1 YJ14_YEAST	P47029 saccharomyc
58	131.5	3.2	1379	1 YFP9_SCHPO	O14066 schizosacch
59	131	3.1	1014	1 HEX2_YEAST	Q00816 saccharomyc
60	131	3.1	1093	1 TMF1_HUMAN	P82094 homo sapien
61	131	3.1	1358	1 SIR4_YEAST	P11878 saccharomyc
62	131	3.1	1612	1 DNM1_PEARL	Q27746 patacentrot
63	130.5	3.1	704	1 MSN2_YEAST	P33748 saccharomyc
64	130.5	3.1	1547	1 TOP2_BOMMO	O16140 bombyx mori
65	130.5	3.1	1449	1 IGA4_HAEIN	P45386 haemophilus
66	130.5	3.1	2805	1 MABA_HUMAN	P78859 homo sapien
67	130	3.1	969	1 SABC_STRSL	O55242 streptococ
68	130	3.1	1283	1 OSH2_YEAST	Q12451 saccharomyc
69	130	3.1	1829	1 RML_MOUSE	O70472 mus musculu
70	130	3.1	6669	1 NEBU_HUMAN	P20929 homo sapien
71	129.5	3.1	1001	1 RPCR_MOUSE	Q910X5 mus musculu
72	129	3.1	646	1 SGI_BOVIN	P23389 bos taurus
73	129	3.1	1026	1 STAU1_DROME	P25159 drosophila
74	129	3.1	1031	1 YD69_SCHPO	Q10496 schizosacch
75	129	3.1	1570	1 P3K1_DICD1	P54673 dictyostell
76	129	3.1	2278	1 FABI_YEAST	P34756 saccharomyc
77	129	3.1	3038	1 TRIO_HUMAN	O75662 homo sapien
78	128.5	3.1	1137	1 MSB1_YEAST	P21339 saccharomyc
79	128	3.1	742	1 CD44_HUMAN	P16070 h cd44 anti
80	128	3.1	817	1 YG4A_YEAST	P46949 saccharomyc
81	128	3.1	1039	1 MSU1_DROME	P50535 drosophila
82	128	3.1	1616	1 P200_MYCGE	Q49429 mycoplasma
83	127.5	3.1	782	1 Y044_UREPA	O97341 ureaplasma
84	127.5	3.1	954	1 B1R1_YEAST	P47134 saccharomyc
85	127.5	3.1	1334	1 J1P3_HUMAN	Q94966 homo sapien
86	127.5	3.1	1466	1 SPA2_YEAST	P21301 saccharomyc
87	127.5	3.1	1597	1 GTF1_STRDO	P11001 streptococ
88	127.5	3.1	1608	1 HLVA_SERMA	P13520 serratia ma
89	127.5	3.1	1701	1 MSP1_PLAIVB	P18199 plasmodium
90	127.5	3.1	2103	1 F8A_PIG	P08869 plasmodium
91	127.5	3.0	1167	1 CAGA_HELPJ	P12263 sus scrofa
92	126.5	3.0	1816	1 LMA4_MOUSE	Q92161 helicobacte
93	126.5	3.0	2415	1 SPCA_MOUSE	P13395 mus musculu
94	126.5	3.0	3562	1 PGCV_CHICK	O90953 gallus galli
95	126.5	3.0	918	1 YMB_CABEL	P34487 caenorhabdi
96	126	3.0	1132	1 YKX5_YEAST	P34250 saccharomyc
97	126	3.0	1228	1 ATC_PLAIVB	Q00853 plasmodium
98	126	3.0	1455	1 GTFC_STRMU	P13370 streptococ
99	126	3.0	3060	1 BBPB_HUMAN	O8wX68 homo sapien
100	126	3.0			

ALIGNMENTS

RESULT 1

```

YF16_YEAST      STANDARD;      PRT;      1233 AA.
ID   YF16_YEAST
AC   P43537.1
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   01-NOV-1995 (Rel. 32, Last annotation update)
DE   Hypothetical 137.7 kDa protein in UGSI-FAB1 intergenic region.
GN   YF016C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetaceae; Saccharomycetaceae; Saccharomycetes.
OX   NCBI_TaxID=4932;
RN   11
RP   SEQUENCE FROM N.A.
RC   STRAIN=5288C / AB972;
RX   MEDLINE=95400292; PubMed=7670463;
RA   Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA   Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA   Yamazaki M., Tashiro H., Eki T.;
RT   "Analysis of the nucleotide sequence of chromosome VI from
RT   Saccharomyces cerevisiae";
RL   Nat. Genet. 10:261-268(1995).
CC   -1- SIMILARITY: SOME, TO MAMMALIAN NEUROFILAMENT TRIPLET M PROTEIN.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
CC   EMBL, D50617; BAA09255.1; -
CC   DR   PIR; S56271; S56271.
CC   DR   SGD; S0001912; YF016C.
CC   KW   Hypothetical protein.
CC   SQ   SEQUENCE 1233 AA; 137697 MW; C8A7CD2C6F0892F6 CRC64;

Query Match      4.1%; Score 171; DB 1; Length 1233;
Best Local Similarity 19.8%; Pred. No. 0.12;
Matches 189; Conservative 142; Mismatches 402; Indels 222; Gaps 42;

QY 12 VENNRRVYIDGKATQKTEMLTPDEVSREGINAQIYI-----KITD-OGYVTS 61
DB 151 IKETSTNNVABG-----TENVP--IKESTGLEVGNBPIRRKKKKKTTNRGRSS 202
QY 62 HGDHYHYNGKVPYDAII--SEELLMKDPNYLKDEDI-----VNEVGKGYIVKD 110
DB 203 NPADTTDSKOSTLDSILVGIIEVLOEDGS--KNEDIKVINIVODEPVNVEKMDIRTRNE 259
QY 111 GKYYVYLKDAAHADNV-----RTKEINRQKQESHQHEGG--TRNDGAVALLASQ-- 160
DB 260 SSDKTFDIDVPRKMDVDETSSENNINEEKAHEHTLPREEMELIYNVNGNNAASFHQLEP 319
QY 161 -GRYTTDDGYLFNADIEDTGDAIVPHGDHYHYPKNELSASELAAEAFLSGRLS 219
DB 320 HGLENGDE---NGASTKRDVSESLTKNGFFKNESKHLAKGKQOTE---SRDGIS 372
QY 220 NSRTYRRQNSDNTSRTNWVPSVSNP-----GTTNTNTSNNSTNSQASQSNDI----- 267
DB 373 PSVLAKNQKETEIGKEDHVFEGKDEKCKRELSVNHENMNSHFNAGSSDIIIPETE 432
QY 268 -----DSLKLQYLKPLSORHV-----ESGGLVFPDPAQ--ITSR 299
DB 433 RETYDDEMTGPTKRTISDNKKNLQHGNTNDISVEVEKEEKEEENSTFSKYKKNVYGE 492
QY 300 -----TARGVAVPHGDHYHPIFYQSEMLERIRIRIILPLRYSN----- 338
DB 493 QEAVANNVSGTEESTSKGEIIMGD-----EKQSDAGEK-SSIIIEBSASASAKIS 544
QY 339 -HWPDRPPOPSQPTPEPSPGQOPAPNLKIDSNSSLVSQLVAKVGEVGFEEKGISR 396
DB 545 KDNVLVLEDEAEAPTEQENKPTFVVGIEDIDPADPRD-----VEIIVAEKNIIPEDLEVAK 599

```

```

QY 397 -----YVPAKDLPEETVKNLESKLSKQESVHTLTAKKENVAPR-----DOEFYDK 442
DB 600 EDQGEBOVKLDEPVPYAMKDDKIATKAGBSISDEKKKQEGTAELSNKAKKEVDETARIS 659
QY 443 AYNLTEAHKALFYNNK-----GRNSDFQALDKILRELNDSTNKEKLVDDLAFLLAPI 495
DB 660 AEGVVEKSKTPESPKVKVRCSTSRPDLQINERDPEVLKDVAVPDEVDVPELATIEN 719
QY 496 THERPLGKPNISQIETEDDEVIRIADLADKYTTSDDGIYIDEHD-----IISDEGDVYVPHM 550
DB 720 SEEDPKSQORVQISTEQAFETIQKDMGDVGSYTS--FKEBKPRKFEITQGDGKITGKDT 776
QY 551 GSHMIGKDSLSDKEXVAAQAVTKEKGLPSS-----PDADVKNPNTSDSAAATYNR 602
DB 777 NHEGAEVTAELNKENDVAVATSKEDIEFKCEPAPETPIED--GTCTEAEVSKDAE 830
QY 603 VKGSKRIPLYALPVMVEHTVEVKNKNLI-----PHKDHYNIKFAMFPDHTY 650
DB 831 VKAE--LENIDAPKAEVTAELNKENDVAVTDEDAEVENSEKTEFIKYK--AELGNL 885
QY 651 KAPN--GYTF-----DLFATIKYVE--HPDRPHSNDGNGNASHVLGKGD-- 694
DB 886 DAPKEAEVTAELNKENDVAVATSKEDIEFKCEPAPETPIED--GTCTEAEVSKDAE 942
QY 695 -HSEDPNKPKAEPEPVEETPAPEVPQVE-----TEKYE-AQLKAEVLLAKVTOSL 746
DB 943 AVTKEDEMNENSKIAELKADKDTGQELDDINISDEFQRTVELPLEKQDIDKNGDEDEL 1002
QY 747 KANATETLAGL-----RNNLTLQIMDNNSIMAEKLLALKGSNPSVSKKIN 796
DB 1003 EEEETKEKETSPLDLVVEENIT--EKKNEIKQEEE-VSQDPEFETSISKEAN 1053

RESULT 2
APX_XENLA      STANDARD;      PRT;      1420 AA.
ID   APX_XENLA
AC   O01613.1
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   15-JUN-1998 (Rel. 36, Last annotation update)
DE   Apical protein (APX).
GN   APX.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   11
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=93107151; PubMed=1334959;
RA   Staub O., Verrey F., Kleyman T.R., Benos D.J., Rossier B.C.,
RA   Kraehenbuhl J.-P.;
RT   "Primary structure of an apical protein from Xenopus laevis that
RT   participates in amiloride-sensitive sodium channel activity.";
RL   J. Cell Biol. 119:1497-1506(1992).
CC   -1- FUNCTION: IS PART OF A MULTIMERIC COMPLEX WHICH IS INVOLVED IN
CC   ANILORIDE-SENSITIVE SODIUM CHANNEL ACTIVITY.
CC   -1- SUBCELLULAR LOCATION: Membrane-associated.
CC   -1- TISSUE SPECIFICITY: KIDNEY, PROXIMAL INTESTINE, OOCYTES, AND
CC   TO A LESSER EXTENT IN THE DISTAL INTESTINE, STOMACH AND EYE.
CC   -1- SIMILARITY: SOME, TO HUMAN APXL.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
CC   EMBL, Z14997; CAA78718.1; -

```

```

DR PIR; A44361; A44361.
KW Membrane; Sodium transport; Transport.
FT DOMAIN 111 114 POLY-SER.
FT DOMAIN 559 564 POLY-SER.
FT DOMAIN 1048 1051 POLY-THR.
SQ SEQUENCE 1420 AA; 159467 MW; 0DD8B5C1413FFB6C CRC64;

Query Match
Best Local Similarity 19.8%; Score 161; DB 1; Length 1420;
Matches 180; Conservative 111; Mismatches 318; Indels 302; Gaps 44;

QY 42 EGINAEQVIYIKITDQGYTSHGDY-----HYNGKVPY-----DAISRE 82
DB 26 ERISPVRMTTLVDSAYSSFSGSSYVPEYONSFOHDGCHYNEQJLSYMDSEYVRAIYNPS 85
QY 83 LLMKDPNYKTKDEDIYNEVKGGVIVKDGK-----YYVYLKDAHANVRIK--- 129
DB 86 LLDKDGCVY---NDIVSE-HGSSKVALSGRSSSSSLCSDDNTTSVHRTSPAKLDNYVTNLD 140
QY 130 -----EELN-ROKQESHQREGGTPRND--GVALARSGR-----YTTDDGYIFNA 173
DB 141 EKNIVGDPINMHKQKQRPNHKAYGLORNSPTGINSLOEKENQLYNSNMEIKDNVFGS 200
QY 174 SPITIEDTGVAVVPHGDYH---YIPKNELSASEIAAEAPLISGRGLSNSR---TYR 225
DB 201 LDVLQADGD---IMQDSYTONALYFPQNOQOYR---NTQYPGANRMSKEQFKNVDVQ 253
QY 226 RQNSDNTSRTNWVPSVSNPGTNTNTNNSNNSQASQSDNDLSLKQYKPLPSGRHVE 285
DB 254 KSENEETEDG--PYLTQDQGF-----VQGYASDVRTSPKNI-----RRSLK 294
QY 286 SDGLVDFPQITSRTARG--VAVPHGDYHF-----IPYQSELEERARIIT--- 331
DB 295 KSA-----SKTIYAHDSQGSQWIMKPKQDTPSPFSEGTITDMYDNRQDIDKRSIUSTA 350
QY 332 --PLRYRSHWVPDSRPEQSPQPTPEPSPGPAPNLKIDSSNLVSQLVKRVGGEYVF 389
DB 351 SOSLYESNEVDY-----SGPPLKAMNSKNEVDQTLG----- 381
QY 390 EERGISRYVAFKDLPEETVKNLESKLSKQESVSHITLAKENAVAPDQEFYDKA-----Y 444
DB 382 -----FQKDAIVKSIPLLSQQLQOQKCKSHPL-----SDLNCKIKITASTPMLY 425
QY 445 NILTEAHKALF-----XNKRNSDFOALPKLLERLNDESTNKEKLVDDLAFLAPITHE 499
DB 426 HLAGGRHSFPIAVNHTNPAQOEKLEKLEKTERNNNIS-----VLQISEPPDNH 476
QY 500 RLKQNSQIEYTEDEVRIQADAKYTTSDGYIFDEHDIISDEGDVAVTPHMGSHWIGKD 559
DB 477 KLPKMS-----LTQLADLHDSVSG-----GNSGNLNSABE 508
QY 560 SIS---DKEVAAQAYTEKEGI---LPPSPADYKANP----- 591
DB 509 SLMDNYIEKLVQAQKVKLRETSFKRKDLQMSLPCRFKMLNPKRPRTIDHFRSYSSSANE 568
QY 592 -----TGDSAAIY-----NRVKEGKRIPLVRLVYWEHTVAVNGNMLIPIHK-D 635
DB 566 SAYLOTQNSADSSYKQDTEKAVATRIIGRKRI-----TEQKKLCYSEBEKD 617
QY 636 H--YHNIRKFWPDHTYKAPNGYTTLEDLPAITIKYVVEHPDERPHSNDGNGNASEHVLGKK 693
DB 618 HIGIQKSNFAWKEEPLFANRREMSDDISANRIKYLE-SKERNNSS---SNLSKTLKQI 673
QY 694 DH-----SEDPNKFKADEPVEETPA-----EPEVPQVETEKVEAOLK--- 732
DB 674 QHNALVQYMERKTQNPNSN---POVOMERTSLGLPNYEWMSIYSETSSSDASQCYLRR 730
QY 733 -----EAEVLAKVT-----DSSLKANATEITLAGRNLLTQIMNNSIMAEKTLML 781
DB 731 RBAQSSSDAIVATVNNDRFGKTSPLGRSAEKTAGVQK---TFSDQRTLDGSGEH--- 783
QY 782 LKQSNPSSVSK 792

```

```

DB 784 LEGSPSSLSQK 794

RESULT 3
ID UB8_HUMAN STANDARD; PRT; 1118 AA.
AC P40818;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Ubiquitin carboxyl-terminal hydrolase 8 (EC 3.1.2.15) (Ubiquitin
DE thiolesterase 8) (Ubiquitin-specific processing protease 8)
DE (Deubiquitinating enzyme 8).
GN USP8 OR KIAA0055.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=bone marrow;
RX MEDLINE=96051398; PubMed=7584044;
RA Nomura N., Nagase T., Miyajima N., Sazuka T., Tanaka A., Sato S.,
RA Seki N., Kawarabayashi Y., Ishikawa K.-I., Tabata S.;
RT "Prediction of the coding sequences of unidentified human genes. II.
RT The coding sequences of 40 new genes (KIAA0041-KIAA0080) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 1:223-229(1994).
CC -I- CATALYTIC ACTIVITY: Ubiquitin C-terminal thiolester + H(2)O =
CC ubiquitin + a thiol.
CC -I- SIMILARITY: Belongs to peptidase family C19.
CC -I- SIMILARITY: Contains 1 rhodanese domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; D29956; BAA06225.1; -.
CC DR MEROPS; C19.011; -.
CC DR Genew; HGNC:12631; USP8.
CC MIM; 603158; -.
CC DR GO; GO:0004197; F:cyteine-type endopeptidase activity; TAS.
CC DR GO; GO:0004843; F:ubiquitin-specific protease activity; TAS.
CC DR GO; GO:0008283; P:cell proliferation; TAS.
CC DR GO; GO:0007048; P:oncogenesis; TAS.
CC DR InterPro; IPR001763; Rhodanese-like.
CC DR InterPro; IPR001394; UCH-2.
CC DR Pfam; PF00581; Rhodanese; 1.
CC DR Pfam; PF00443; UCH; 1.
CC DR SMART; SM00450; RHOD; 1.
CC DR PROSITE; PS50206; RHODANES_3; 1.
CC DR PROSITE; PS00972; UCH_2_1; 1.
CC DR PROSITE; PS00973; UCH_2_2; 1.
CC DR PROSITE; PS50235; UCH_2_3; 1.
CC DR Ub1 conjugation pathway; Hydrolase; Thiol protease; Multigene family.
FT DOMAIN 195 313 RHODANES.
FT ACT_SITE 786 786 BY SIMILARITY.
FT ACT_SITE 1059 1059 BY SIMILARITY.
FT ACT_SITE 1067 1067 BY SIMILARITY.
SQ SEQUENCE 1118 AA; 127523 MW; 8B884B7A842F9A9A CRC64;

Query Match
Best Local Similarity 19.0%; Score 158; DB 1; Length 1118;
Matches 173; Conservative 143; Mismatches 336; Indels 258; Gaps 45;

QY 6 LVQARTVKNRNVSYIDGKQATQKTEMLTPDEVSKREGINAEQVIYIKITDQGYTSHGDH 65
DB 11 LVYSSSLDLNKK-----KTE-VKPEKISTKSYVSHALKIFKTAEECRDLR--DE 56

```

```

OY 66 YHYNGKVPYDAIIEELLMKDPNYKLEKDEDIVNEKGVYIKVDGKYVYLKDA----- 121
DB 57 ERAYVYMKYVTVN--LTKRPDK--OQDYFHLPGNITKKAVERAERLSESLKRY 113
OY 122 HADNVRK--EINRQKQ--EHSQREGGTPRNDGVAL----- 156
DB 114 EEAVERKLEEKDROEABORLOQKQOETGREDGTLAKSLLENVLDSKDTOKSGEKNE 173
OY 157 ---ARSQRYTTDDGIF---NASDI-----EDTGDAIVPHGDVHYIIPKHELAS 203
DB 174 KCETEKATITAKELYTMATDKNISLIIMDARRMODYDSCIL---HSLVPEEAIISPG 229
OY 204 ELAA--AEAFISGRGINSRTRYRONSNTSRTNVPSVN--PSTTNNTSNNNTNSQ 260
DB 230 VTASIEIHLR---DDSKDTMKKRGVNEVYVLLDFSSAKDQITITRLSK----- 278
OY 261 ASQSDNIDLKQLYKLPLSQRHVESDG---LVDPQAQTSRTAGVAVPHGDVHYFI 315
DB 279 ---DALFKWESKTVLRNEPLVLEGYENMLLCYQYTTNAK---VTPP----- 320
OY 316 PYSQMSLEERLARIIPLRYSNMHWPPSRPEQSPQTPPESSPQAPANIKIDSNSL 375
DB 321 PRRO---NEEVSISLDFTPS---LEBSIPSKPAQO-----TPPASIEVDENIEL 364
OY 376 VSQVLRKVG-----EGYVFEKGISRYVFAKOLPSETVKNL-----ES 413
DB 365 ISGNERMGPLNISTPVEPVASKSDVPII---QVPSTINQIOTRTKKPAVKLPEE 420
OY 414 KLSKQESVSHLTJAKKENAVPRDOFPYDKAVN---LLTEAHK-----ALEXNGRNSD 463
DB 421 HRIKESINHEQSQSQSGKVIIPDRSTKPVFSPTMLTDEKARIIHAETALLMEKQGE- 479
OY 464 FOALDKLERLNDSESTNKEKLVDDLAFAPITHEERLGRKPNISQI---EYTEDVRILQOL 520
DB 480 ---KELEROQEE---QKEKLKKE-----EQEQAKKKQKQAEENEITEKQOKAEE 524
OY 521 ADKYTTSQGYIFDEHDIISDEGDVATPHMGSHWIGKDSLSDKEKVAQAQATKEKGLP 580
DB 525 MEKESSEQAKKEDK--TSARKGKEIT-----GVKQSKSEHTSDAKSVERGRKRC 575
OY 581 PSPDADVKA-----NPTGDSAAAIYNRKGEKRIPLVRLPYWVEHTEVEVNGNL----- 629
DB 576 PTEBIOKSTGDPHTSVTGDSGSKPFKIQQPSGILRTGTFREDIDTGERNAKQREP 635
OY 630 -----IIPKHDHYNIKFAMFDDHYTAAANGT---LEDI-PATIKTY-----VEH 671
DB 636 LTRARSEEMGRIVP-----GLPSGWAFLDPIITGTFRRYHSPNTNVH 677
OY 672 ---PDERPHSNDGNGNASEHVLGKKHSEDENKQFKADEBP-----VEETPAPE 718
DB 678 MYRPEMASSAPSPSTPPTHAKAKFOIAPARD--REPSKLKRSYSSPDITQALOEKKRKPT 735
OY 719 V-POVETEKVEAQLKEAEVLLAKVTDSIK-----ANATETTLAQLRN-----NLTLQ 764
DB 726 VTPVTRNENKPTCYPKAEI--SRLSASQIRNLNIPFGSGPALITGLRNLGNTCYMNSILQ 793
OY 765 IMDNNSIMAE 774
DB 794 CLCNAPHLAD 803

```

```

RESULT 4
DNM MYCPN
ID DNM MYCPN STANDARD; PRT; 910 AA.
AC P75354;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DnaA-like protein MG200 homolog.
GN MPN19 OR MP035.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
NCBI_TaxID=2104;

```

```

RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8946633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RT Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RL pneumoniae."
CC Nucleic Acids Res. 24:4420-4449(1996).
CC -I- SIMILARITY: Contains 1 J domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB000004; AAB95683.1; -.
DR PIR; S73361; S73361.
DR HSSP; P23685; 1HDJ.
DR InterPro; IPR001623; DnaJ_N.
DR InterPro; IPR003095; Hsp_DnaJ.
DR Pfam; PF00226; DnaJ_1.
DR PRINTS; PRO0625; DNAJPROTEIN.
DR SMART; SM00271; DnaJ_1.
DR PROSITE; PS00636; DnaJ_1; 1.
DR PROSITE; PS50076; DnaJ_2; 1.
KW Hypothetical protein; Chapterone; Complete proteome.
FT DOMAIN 4 73 J-DOMAIN.
SQ SEQUENCE 910 AA; 100190 MW; 125D0E37D20221A7 CRC64;

Query Match 3.8%; Score 157; DB 1; Length 910;
Best Local Similarity 19.0%; Pred. No. 0.43;
Matches 161; Conservative 127; Mismatches 305; Indels 254; Gaps 44;

```

```

OY 66 YHYNGKVPYDAIIEELLMKDPNYKLEKDEDIVNEKGVYIKVDGKYVYLKDA----- 114
DB 34 YHPDRNAPPAQAQFAFINEANDVLSNPKRANRDKYGHQVDN-BGQFAQADVPSFF 92
OY 115 VYLKDAHADNVRITKEINRQKQESQHREGGTPRNDGVALAASQGRYTTDDGYIFNAS 174
DB 93 EIEKSGAFDL---SSNTKKKEKTKKKGW-----FWGSKQESTSDTTEY--A 140
OY 175 DIEDTGDAIVPHGDVHYIIPKHELASLAAAEALSGRNTSNRTTRONSNTSR 234
DB 141 DV--DAGLEDPVPPQSDYPPDIIPVDARIEEVDQ--AYADDIPVDAGMDWEQNAEVANSA 197
OY 235 TNWVPSVSNQCTTNTSNNSTNSQASOSNDIDSL----- 271
DB 198 SEIIPDV-----DAGLADENITSSAAPQASDWEAMTGNBEYGFDAAGENWKGPFDEA 251
OY 272 -KQLYKLPLSQRHVESDGLVDPQAQTSRTAGVAVPHGDVHYFI PYSQMSLEERLARI 330
DB 252 GQWVLEETEPSSVSNQETTTSDAVTAAT-----VEETQ----- 288
OY 331 IPLRYSNMHWPPSRPE-----QPSQPTP-----EPSPGP-----QAPNLIK 369
DB 289 ---DSWTANAPAPVDEPVELOPTEPEPIITLSEPEAPASVIEPTPEIEE 341
OY 370 DSNGLSVQVLRKKGEGYVEFEKGISRYVFAKOLPSETVKNLEKLSKQESVSHLTAKK 429
DB 342 TTSAVENDASVK---ADVSDADATPEPTQDTISPEQ--ETDAALAEINHT-TADL 394
OY 430 ENVAPRDOEFYDKAYNLTTE--AHKALFXNKGSRSDFOALDKLERLNDSESTNKEKLVDD 487
DB 395 E---PAEV---SKTNLEQDVVAKNFSSEESTVDTAAIDPVVEQATETSTNGFKFP-N 446
OY 488 LLAFL-----APITPHERLGRKPNISQI---EYTEDVRILQALA----- 521
DB 447 FSSFVLSQDNPNQPTPHHEEDAAAPPTVDETSGESTAADEVTAESTVELETAAEIINP 506

```



```

RP SEQUENCE FROM N.A.
RX MEDLINE=96017704; PubMed=7593161;
RA Espenade P., Gismo R.E., Holzmaier E., Teung P., Kaiser C.A.;
RT "Yeast Sec16 gene encodes a multicomponent vesicle coat protein that
RL J. Cell Biol. 131:311-324 (1995).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=SS288C / AB972;
RX MEDLINE=97313271; PubMed=9169875;
RA Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansgore W.,
RA Azeiteiro R., Aparicio A., Barrett B.G., Badcock K., Bensy J.M.,
RA Bocteanu D., Bowman S., Bruckner M., Carpenter J., Cherry J.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Dillus H., Dipolo T., Dubois E., Duesenrodt A.,
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Hebling U., Heumann K., Hilbert H., Haller L.,
RA Hunkeler-Smit S., Hyman R., Johnston M., Kaiman S., Klatte K.,
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marathe R., Messenguy F., Mewes H.-W., Mitrakopoulou S., Moestl D.,
RA Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA Petzel F.X., Pohl T.M., Purnelle D., Schater M., Scharte M.,
RA Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tectelin H.,
RA Ureastetaru L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA Zhong W.V., Zollner A., Vo D.H., Han J.;
RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI."
RL Nature 387:103-105 (1997).
CC -1- FUNCTION: INVOLVED IN THE BUDDING OF TRANSPORT VESICLE FROM THE
CC ENDOPLASMIC RETICULUM. THE C-TERMINAL INTERACTS WITH SEC23 AND
CC WITH THE CYTOSOLIC DOMAIN OF SED4. COULD THEREFORE BE A
CC CONSTITUENT OF COP1 VESICLE COAT. N-TERMINAL OVEREXPRESSION
CC CAUSES A LETHAL SECRETION DEFECT.
CC -1- SUBCELLULAR LOCATION: ON THE ENDOPLASMIC RETICULUM AND ON VESICLES
CC WHICH BUD FROM IT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, U23819; AAC49088.1; -
DR EMBL, U41849; AAB68254.1; -
DR PIR, S61103; S61103.
DR SCD, S0006006; SEC16.
KW Transport; Protein transport; Golgi stack; Endoplasmic reticulum.
FT DOMAIN 1997 2094
FT MUTAGEN 1059 1059 L->S; IN SEC16-4; TS ACCUMULATION OF ER
FT 1084 1084 MEMBRANES.
FT MUTAGEN 1084 1084 L->P; IN SEC16-3; TS ACCUMULATION OF ER
FT 1089 1089 MEMBRANES.
FT MUTAGEN 1231 1231 L->P; IN SEC16-2; TS ACCUMULATION OF ER
FT 1231 1231 MEMBRANES.
FT MUTAGEN 1231 1231 W->R; IN SEC16-1; TS ACCUMULATION OF ER
FT 522 522 MEMBRANES.
FT CONFLICT 522 522 MISSING (IN REF. 1).
FT CONFLICT 560 560 I -> F (IN REF. 1).
SQ SEQUENCE 2195 AA; 241694 MW; 757B7A7231BEE6F0 CRC64;
Query Match 3.8%; Score 157; DB 1; Length 2195;
Best Local Similarity 19.7%; Pred. No. 1.4;
Matches 124; Conservative 109; Mismatches 241; Indels 156; Gaps 31;
OY 188 HGDHNYPKHLELSELAALAEFLSGKGNLSRTYRQNSDNRTRNVPVNPPT 247
DB 29 HSELELPBSTINS-----FNDSVNRKTESDLSKSD-VPPVS-----S 68
OY 248 NTNTS-----NNTNSQASQSDNDI-----DSLKLQYLKPL 279
DB 69 STNISPAVETQLEPTQELNHLKLNLDSDQHDITADSNLDLPNLSIVEDSVITQI-KRAM 127

```

```

OY 280 SORHVESDGLVDPDAOTSR-----TARGVAVPHGDHYHFIYSGOMSE----- 322
DB 128 SQEYET-----AAHLSRNPISLDVYAGELHNNHETQKIAVAVEDSNEEGEWH 181
OY 323 -LEERIAARIPLRYRSNHWVDSRPPSPQPTPEPSGPQAPNLKIDSNSLSVLQV 380
DB 182 SIILSLNDATPSQY--NHFLPSD-----GNLSPELSSGDTPTHNVPLGTQDN----- 228
OY 381 RKVGEVFEKGRISRYFAD-LPSETVKLESLKSQESVSHTLTRAKENVA-PRODF 439
DB 229 -EINDEYCDNKEIS--LNANNVLPDELSEDERLKLETHVS--TEEKQDIA--DET 281
OY 440 YDKAVNLTLEAKLIPXKNGNSD--FOADKLRLERLNDSTNEKLVDDLLAFAPITH 497
DB 282 AENLFTSTPSEKKNISGDTSMLO-----DDESDQVPMEEVAKPF-----H 328
OY 498 PERLGRNPQIETEDEV-----IAQLADKYTTSQGYTFDE--HDIISDEGDAVTP 548
DB 329 NENTNNTQESAPNTDDDKQEGNEALKKSSECTAABERSYSEETSEDIFGHDKQVVE- 387
OY 549 HMGSHWIKGSLSDKEKVAQAYTKKGLIPPEPDV--KANPTGSAALINRVGE 606
DB 388 --GNDPTGKNIENESQKMGEGNHK---LPLSAEADIIPGKXIDQDAEDLFTQSSGD 441
OY 607 KRPLVRLPYVVEHTVEKNGNLIIPKDHVHNI-----KFAW-----FDDHTYKAPN 654
DB 442 -----LGEVLPEWSTDKNADVTSKOEKHEDLPAASGNDEKLPWEVSDEGVSSGKTEN 494
OY 655 G-YTLIEDLFAITKY-YVEHPDERPHSNDGMSAEHVLGKKHSDPNKFADEPVEE 712
DB 495 SMQSTELIAHQKSPLENDLDDDDSLFLASE-----BEDTVPTNDTNTNLSPVEE 550
OY 713 TPAREVPQVTEKVEKQKAEVYLAKVT 742
DB 551 KKASRYKPIIEE---EAGMRQEVHFTNTT 577
RESULT 7
MAPX DROME STANDARD; PRT; 1185 AA.
AC P23226; Q9V9S1;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE 205 kDa microtubule-associated protein.
GN MAP205 OR CG1483.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RN SEQUENCE FROM N.A.
RP MEDLINE=9115949; PubMed=1703540;
RX Irminger-Pinger I., Laymon R.A., Goldstein L.S.B.;
RT "Analysis of the primary sequence and microtubule-binding region of
RL J. Cell Biol. 111:2563-2572 (1990).
RN [2]
RN SEQUENCE FROM N.A. (ISOFORM B3).
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=1073113;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang O., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle G., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

```

QY	122	HANNVTKKEI--NRQGO-----EHSOREGCTPRNDGAVVLARSQGRVTTDDGIYF	17
Db	4	HEMNAQDNDVYLQNRNLAEISLQICGAGGHNPHLADATGGNGCAPGIAPSK-----	52
QY	172	NASDIIEDTGDAYIVPHGDHYHYPKNELSASELAAAE-----AFLSGRGLSNS	221
Db	53	--SDEVGBED-----EEMKYI--HEVROSEKYLQOEKPLUTKETGNGFQGR-----	95
QY	222	RITRRQNSDNTSRITW-----VPSVSNPCTTNTNTSNSNTNS-----QASQND	266
Db	96	-----DSNOVHGGAAVFNLVEDEVEVINKNDPFSNTNNTTSTDEVARQAQOEPNQ	149
QY	267	IDSLIKOLYKPLPSORHVESDGLVEPDPQIISRTARGAVALPHG-----DHYHF	311
Db	150	LPRQLDQ-----OOQISQGVHEDPROEDEHSSVATTYGTSSISENNSSPLDOEEV	207
QY	315	IPYSQ-----MSELEBRIARIIPLRYRSHHWVPSRPEPSPOTPE	356
Db	203	VMAVQIVGQEQVLDPFNKENSIFYKNLBEHNSQLN---NAVAFVPGVGSQSSPLPAE	25
QY	357	-PSPGQAPNLKIDSNSSLSQVLRXKYGEGVFEKISRYVAKDLPSTVKNLESKL	415
Db	260	DPLPGVQPPPLPGCTGLDLVAESPRK--EPARIMDDIA-----VPDEREPDIADM	310
QY	416	SKQESVHTLTAKKEVAPRDEPFQKAYNLLTEAHKLPXNKRNSPQALDLEBRLN	475
Db	311	R-----PHELEQSDTFGAGHLEMQ-----LLNGIGTDAQALRPVLDHGR	351
QY	476	DESTNKEKLVDDLFLAFIATHTPERLKGNSQIE-----YTEDEVRIAQLADKY	524
Db	352	ETSVDMW-LPLDQVPNDADIMKQSIYAEHNSIEDILNSVQPLQTCDDKELIHVEKE	410
QY	525	TTSDGVIIPBHDILSPBGAYVTPPMGSHWTKGDSLSDKKVA-----AQATKE	578
Db	411	HVSKSPSTBELPQSD-----FPNNOESHTLFNNTEODPMQASFLYEHTSOKAQEGQF	46
QY	576	KGLIPSPADYKANT--GDSAAIIVRVKGEKIPLVRL-----	614
Db	465	QMOVL-PAECSDIFADQSLDLDTSAF---QLSEADSPAKLELSQQAQIVDTPSPSS	522
QY	615	---PYVNEHTVEVKGNLIIIPKHQYHNIKFAMFDDHTYKAP---NGYLTEDLFATIKY	667
Db	521	TAEKHLVEDTKEIVEEYTLDP--ESHFPGVV-----SSQAPQLQFGKHTLPSIHSCKH	572
QY	668	YV---EHPDRPRHSNDGWSGASHVLCKGDHSDPKYKAKADEPVEETPAPREVPQVETE	722
Db	573	RVASEQND--ENAVFESVSGYETQNFDEISSPPEGINPRAQRP--TPRALVIEQANTM	622
QY	726	KVAA-----QLKEAEVLAKVTDS--LKANATETTLGLRNNLTLLQIMD-----	767
Db	628	MEVGGMPPIPASEDFALCDKVAKSNSNEVEDHRSQQAFAVEELLHPGVDAVQVENLGT	687
QY	768	NNSIMAEKLLALLKGSNP-SSVSKEKI	795
Db	688	EKNFVVEEBRLPLISVDEIPLISSAKSEKXL	716
RESULT 8			
ANK2_HUMAN			
ID	ANK2_HUMAN	STANDARD;	PRT; 3924 AA.
AC	Q01484; Q01485;		
DT	01-APR-1993 (Rel. 25, Created)		
DT	01-OCT-1996 (Rel. 34, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Ankyrin 2 (brain ankyrin) (Ankyrin B) (Ankyrin, nonerythroid).		
GN	ANK2.		
OS	Homo sapiens (human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
OX	NCBI_TaxID=9606;		
RF	[1]		
	SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).		

CC TISSUE=Brain stem; PubMed=1830053;
 RX MEDLINE=9130246; PubMed=1830053;
 RA Otto E., Kunimoto M., McLaughlin T., Bennett V.;
 RT "Isolation and characterization of cDNAs encoding human brain
 RT ankyrin reveal a family of alternatively spliced genes.";
 RL J. Cell Biol. 114:241-253(1991).
 RN [2]
 RP REVISIONS.
 RA Carpenter S.;
 RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 1).
 RC TISSUE=Brain stem;
 RX MEDLINE=94075409; PubMed=8253844;
 RA Chan W., Kordeli E., Bennett V.;
 RT "440-kD ankyrinB: structure of the major developmentally regulated
 RT domain and selective localization in unmyelinated axons.";
 RL J. Cell Biol. 123:1463-1473(1993).
 RN [4]
 RP SEQUENCE OF 463-495 FROM N.A.
 RX MEDLINE=92009921; PubMed=1833308;
 RA Tse W.T., Menninger J.C., Yang-Feng T.L., Franke U., Sahr K.E.,
 RA Lux S.E., Ward D.C., Forget B.G.;
 RT "Isolation and chromosomal localization of a novel nonerythroid
 RT ankyrin gene.";
 RL Genomics 10:858-866(1991).
 CC -I- FUNCTION: Attach integral membrane proteins to cytoskeletal
 CC elements. Also bind to cytoskeletal proteins.
 CC -I- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=3;
 CC Name=1;
 CC IsoId=Q01484-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q01484-2; Sequence=VSP_000267, VSP_000268;
 CC Name=3;
 CC IsoId=Q01484-3; Sequence=VSP_000268;
 CC -I- TISSUE SPECIFICITY: PLASMA MEMBRANE OF NEURONS AS WELL AS GLIAL
 CC CELLS THROUGHOUT THE BRAIN.
 CC -I- PTM: PHOSPHORYLATED AT MULTIPLE SITES BY DIFFERENT PROTEIN KINASES
 CC AND EACH PHOSPHORYLATION EVENT REGULATES THE PROTEIN'S STRUCTURE
 CC AND FUNCTION (POTENTIAL).
 CC -I- SIMILARITY: Contains 23 ANK repeats.
 CC -I- SIMILARITY: Contains 1 death domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X56857; CAA40278.1; -;
 DR EMBL; X56958; CAA40279.2; -;
 DR EMBL; Z26634; CAB42644.1; -;
 DR EMBL; M37123; AAB62828.1; -;
 DR PIR; S37431; S37431.
 DR HSSP; P42771; 1DC2.
 DR Genew; HGNC:493; ANK2.
 DR MIM; 106410; -;
 DR InterPro; IPR002110; ANK.
 DR InterPro; IPR000489; Death.
 DR InterPro; IPR00906; ZUS.
 DR Pfam; PF00023; ank; 23.
 DR Pfam; PF00531; death; 1.
 DR Pfam; PF00791; ZUS; 1.
 DR PRINTS; PR01415; ANKYRIN.
 DR SMART; SMO0248; ANK; 22.
 DR SMART; SMO0005; DEATH; 1.
 DR SMART; SMO0218; ZUS; 1.
 DR PROSITE; PS50088; ANK_REPEAT; 20.
 DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.
 DR PROSITE; PS50017; DEATH_DOMAIN; 1.

KW Cytoskeleton; Alternative splicing; Repeat; ANK repeat;
 KW Phosphorylation.
 FT REPEAT 63 92 ANK 1.
 FT REPEAT 96 125 ANK 2.
 FT REPEAT 129 158 ANK 3.
 FT REPEAT 162 191 ANK 4.
 FT REPEAT 193 220 ANK 5.
 FT REPEAT 232 261 ANK 6.
 FT REPEAT 265 294 ANK 7.
 FT REPEAT 298 327 ANK 8.
 FT REPEAT 331 360 ANK 9.
 FT REPEAT 364 393 ANK 10.
 FT REPEAT 397 426 ANK 11.
 FT REPEAT 430 459 ANK 12.
 FT REPEAT 463 492 ANK 13.
 FT REPEAT 496 525 ANK 14.
 FT REPEAT 529 558 ANK 15.
 FT REPEAT 562 591 ANK 16.
 FT REPEAT 595 624 ANK 17.
 FT REPEAT 628 657 ANK 18.
 FT REPEAT 661 690 ANK 19.
 FT REPEAT 694 723 ANK 20.
 FT REPEAT 727 756 ANK 21.
 FT REPEAT 760 789 ANK 22.
 FT REPEAT 793 822 ANK 23.
 FT DOMAIN 1773 1950 REPEAT-RICH REGION.
 FT REPEAT 1773 1784 REPEAT A.
 FT REPEAT 1785 1796 REPEAT A.
 FT REPEAT 1797 1808 REPEAT A.
 FT REPEAT 1809 1820 REPEAT A.
 FT REPEAT 1821 1832 REPEAT A.
 FT REPEAT 1833 1844 REPEAT A.
 FT REPEAT 1845 1856 REPEAT A.
 FT REPEAT 1857 1867 REPEAT A.
 FT REPEAT 1868 1879 REPEAT A.
 FT REPEAT 1880 1891 REPEAT A.
 FT REPEAT 1892 1902 REPEAT A (APPROXIMATE).
 FT REPEAT 1903 1914 REPEAT A.
 FT REPEAT 1915 1926 REPEAT A.
 FT REPEAT 1927 1938 REPEAT A.
 FT REPEAT 1939 1950 REPEAT A.
 FT DOMAIN 3536 3620 DEATH.
 FT VARSPLIC 1039 1039 O -> OFGLKHLPTAPPPLNEGSLVSRILQLGPRTK
 FT (in isoform 2).
 FT VARSPLIC 1444 3528 /FTId=VSP_000267.
 FT Missing (in isoform 2 and isoform 3).
 FT CONFLICT 475 476 /FTId=VSP_000268.
 FT CONFLICT 971 971 GQ -> PE (IN REF. 4).
 FT CONFLICT 3581 3582 I -> S (IN REF. 1).
 FT CONFLICT 3586 3586 QY -> HA (IN REF. 1).
 FT CONFLICT 3586 3586 I -> Y (IN REF. 1).
 SQ SEQUENCE 3924 AA; 430337 MW; 52AC496C48E29D2 CRC64;
 Query Match 3.7%; Score 154.5; DB 1; Length 3924;
 Best local similarity 18.0%; Pred. No. 4.4;
 Matches 164; Conservative 126; Mismatches 222; Indels 387; Gaps 45;
 QY 7 YQATVVENNRVSTYIDKQAKTKENTLTPDEVSRREGINAEQIYIKITDQGVYVSHGDHY 66
 DB 3022 FEARVKEEOK---IFGLMVDROSQGTPTTPARTP-----TEBGTPTSEQNP 3068
 QY 67 HYVKGVPYDAITISEELMKDDEIVNVEKGVYIKVDGKYVYLKDAHADNV 126
 DB 3069 LFQEGKL-----FEMTRSGAIDMTKSY--ADESTHFF----- 3099
 QY 127 RTKEINRQKQENS--QHREGT-----PRNDGAVALLANSQGYTTDDGYIFNASDIE 178
 DB 3100 ----QIGQSRREPTLSBDVKGATGADPLPLETSABSLALSESKEFYDD-----EADLLP 3150
 QY 179 DTGDAIVPHGDHNYHYPKXELASSELAAAEAFISGKGNLSNSTRYRRQNSDN----- 231
 DB 3151 DVSVEEV-----EELPASDAQLSQMGISASTETPTKENVSVGTDLDP 3193


```

OY      757 LRNNLTLOIMDNNSIMAEKEL 778
Db      1245 IE-----MENEKKALEKEL 1259

RESULT 10
Y373 HUMAN STANDARD; PRT; 1539 AA.
ID _Y373 HUMAN
AC O15078;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein KIAA0373.
GN KIAA0373.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AB002371; BAA20828.1; -
DR Hypothetical protein; Coiled coil.
KW DOMAIN 18 1514 COILED COIL (POTENTIAL).
SQ SEQUENCE 1539 AA; 180065 MW; D901314E981BF001 CRC64;

Query Match 3.7%; Score 153.5; DB 1; Length 1539;
Best Local Similarity 19.2%; Pred. No. 1.4;
Matches 200; Conservative 145; Mismatches 334; Indels 361; Gaps 53;

OY      3 ELGLYQARTVKNENSVYIDGKATQKTEN-----LTPDEVSKREGINAEQIVIKITD 55
Db      306 EQALTYVAR-LLEGRRNAKHL--RQTIGSLRPGSGALPLAQOEKFSKTMIGLQNDKLTIM- 361
OY      56 QGYVTSHDDHYHYNGKVPYDAIIS--EELL--MCDP-----NYKLDKDEI-VNEVKG 103
Db      362 QEMKNSQOEHRMEKKTLEMEIKLKGLELLISTLDYGAQKVIWMHMKITELRLQELKL 421
OY      104 GYVIVDKYVVYLLKDAADNVRT-----KEEINFKOEHSHQREGGTPRNDGAVALAR 159
Db      422 NRELVKQKEIKYLNII-SEVERTISSEBEIVQNKFERHQRQAMQRE---VDLERQ 477
OY      160 QGRYTTDDGYFNASDIID--TGDYIVPHGDHYHYPKNEL----- 200
Db      478 LDIFDRQONELINAAQKEEATGS---IP--DPSILPLP-NQLEIALRKIKENIRIILETR 531
OY      201 -----SASELAAAEFLSGRGNLSKRYRRQNSNTRTWVBSVS---NP 244
Db      532 ATCKSLEELKEKESALRLAEQNLISRDKVINELRLRLPATAREKLIHELGRKEMEP 589
OY      245 GTTNT-NTSNNSNTSOSASNDISDLKOLYKPLSQPHVESDGLVDPDAQITSRTARG 303
Db      590 KSHHTLKTAHQTIANMQA-RLNQEKEVLKKYORLLEKAREEO-----RE 632
OY      304 VAVPHGDHYHFLPYVSQMSLEERLARIIPLRYSNMHWVDSRPEQSPQPT----- 354
Db      633 IVKHEEDLHLHRLHLELDQADSLN-----KFKQTAM--DLMKQSPFTVPPTKHFIRLAE 685

```

```

OY      355 PEPSPGPPAPNLKIDNSSLVSQLVR-----KVEGYVFE 390
Db      686 MEQTVAEQD-----DLSSILVYKTKVVSQDLERQREITELKVEFENIKQLQDENHDE 739
OY      391 EKGISRYVFAKDL-----PSETVKNLESKL----- 415
Db      740 VKKVAEYV--EDLKYLLDQSQESQCKSELOAQKANSRAPPTTMRNLVERLKSQALAK 797
OY      416 -SKQESVSHLT-----TAKKE--NVAPRDEPFYDK----- 442
Db      798 EKQKALSRALLELRAEVNTAAERIIISATSQKAEHLNV---QQIVDRHTRELKTOVED 853
OY      443 -AVNLLTEAHKALFXNKR-----NSDFQ--ALDKLT-----ERLND- 477
Db      854 LNEMLL-KLKEALTKSKRNSLTNDLNDLNNELQKKQKAYNKILREKELEIDQENDELKR 912
OY      478 -----STNKEKLVDDLAFIAPITPERLGRPNQSIETYEDEVRIAQDLADK 523
Db      913 QIKRLTSGLOGKPLTDNKKSLIEEL-----QKVKKLENDLEGKVEEVDLKPMEK 963
OY      524 YTTSDGYTFDE-----HDIISDEGDATVTPHMGSHWIGD--SLSDKEKYAA 569
Db      964 NAKBELLRWBEKGKQAKIEGIRNKLKEKEGEVETLTQQLNTL--KDLFAKADKEKLT 1020
OY      570 QAYTKENGILPPSPDADVKANPTGDSAAATYNRKGEKRIPLVRLPYMVEHTVEVKGNL 629
Db      1021 QRLKLTGM-----TVQVUGI-RALESEKELE-----ELKKRNL 1054
OY      630 IIPKHDHYHNKIFAMPDDHTYKAPNGYLTLEDLFATIKYVE--HDPERPHSNDGWNASE 687
Db      1055 DLENDILYMRHQ--LPRDSVEDLHQNRYLOEKHLALBKQFSKDTYKPSI 1106
OY      688 HVLGKQH-----SDPMKNFPADEEPE--ETPAPPEVQVET-----E 725
Db      1107 SGISDDHCOREQELQKENLKLSSENLELKFQLEQANDLRLKNQVRDLKEMCEFLKE 1166
OY      726 KVEAQLKEAEV-----LLAKYTDSSLKANATETLAGLRNNLTLOIM- 766
Db      1167 KAEYORLGHVGRSGRGSKTIPLEKTIIGLMKKYVEKVGQRNN--EQLKXSGILTSEMA 1224
OY      767 ---DNNSIMAEKLIALL 782
Db      1225 NIEQENELKALELEKLAHL 1244

RESULT 11
Y08 YEAST STANDARD; PRT; 1176 AA.
ID _Y08 YEAST
AC Q08236;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 131.4 kDa protein in REX4-ATP19 intergenic region.
GN YOL078W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=97321807; PubMed=9178509;
RA Tzeremia M., Katsoulou C., Alexandraki D.;
RT "Sequence analysis of a 3.2 kb segment from the left arm of yeast
RT chromosome XV reveals eight known genes and ten new open reading
RT frames including homologues of ABC transporters, inositol
RT phosphatases and human expressed sequence tags.";
RL Yeast 13:583-589(1997).
CC -1- SIMILARITY: BELONGS TO THE SINI FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: Z74820; CA939088.1; -
DR EMBL: Z74819; CA939088.1; -
DR PIR: S66771; S66771.
DR SGD: S0005438; YOL078W.
DR GO: GO:0005737; C:cytoplasm; IPI.
DR GO: GO:0001558; P:regulation of cell growth; IPI.
KW Hypothetical protein.
SQ SEQUENCE 1176 AA, 131378 MW, ED2B30D014E5652 CRC64;

Query Match 3.6%; Score 150.5; DB 1; Length 1176;
Best Local Similarity 18.6%; Pred. No. 1.4;

Matches 157; Conservative 118; Mismatches 254; Indels 313; Gaps 41;

QY 104 GYIVKVDGKYVYLKDAAHADNVRKEINRQKQHSQREGGTPRNDGAVA---LARSQ 160
DB 329 GRIFLTNDK---NDGQKSDSLNANKGI-----HDDGSASAGNSVSRDGLTETE 374
QY 161 GRYTDDGIFNADIE--DTGDAYIVPHGDHYHYPKNEISASELAAEFISG--- 214
DB 375 SNNISMESEYINERKDDIDNFDVTSTNI-----NKVSDLGHESTNDGTAV 421
QY 215 ---RGNTSNSTRYRONSND--TSRTNWPVSVPNGTWTNNTSNNTSAS----- 262
DB 422 NMRDSDSRSNSNEPFAQNRDRITPSSYSKSLSEYSEKRSNDSMSGEMSLDS 481
QY 263 --QSDNDLSLKQYKPLPSQR---HVESDGL---VDPDAITSRTANGAVAPHGDHY 312
DB 482 DMQNTNIPS---HSIPMSMKYGIYHGDDSTLNNVFDKAVLTWNSR----- 526
QY 313 HPIFYQMELEERIKARIIPLRIRSNHWVDSPEQSPPTPEBPQOPAPNLIKDSN 372
DB 527 -----HPKRRRDVTISGKEPTSLTSNRKFSVS 554
QY 373 SSVL---SOLVRKAGEGYVEEKGIRYVPAKOLPESTVKNLSKSK--QESVSH----- 423
DB 555 SNTLSRSPILRKHGR-----TSTASSEMKAAPKSDSVLHARRS 556
QY 424 TLTAKENAVP-----RDOFFYDAVNLLEAHKA-----LFNKNKGNSDFOAL 467
DB 597 TLTLDQHDGQSPVSSVHKSSKEGNLIEKTDYLVSKPRAQSLMNFKNKKRTMTNSV 656
QY 468 DKL-----LE-----RLNDEST--NKEKLVDDLAFLADPTPH 498
DB 657 DVLEVESFYCGDKVPVYESMGLIYIQASKKRYRNSFTTKVRSSTIFEVIGFALFLYST 716
QY 499 ERLGRNSQIEYTEDEVRITAGLAD-----KYTTSQGYFDEHDITISDEGDAVVTYHMGH 552
DB 717 EK--KPDN---FEEDGJTEVDISNPNNSFLKTYDEQGEPEEDN----- 754
QY 553 SHWIGK-----DSLSDKEKVAQAAYTEKKG--ILPSP-----DAD--VKANPTG 593
DB 755 ---FGLDKRKSTIQSISDSEVVLCKVDDAKESQNEIETPLPFTGGGLMDASTLNDNSH 811
QY 594 DSAALAYNRKGEKRIPLVRLPVMVHTVEVKGNLLIIPKHDYHNIKFAEPDHTYKAP 653
DB 812 DTTDGTINOLSPFK--PIIG---NEDDIDKTNGSKII-----DVTYVLYVP 851
QY 654 N-----GYTLEDFAPIIKYVVEHPRPHSNDQMGNASEHVLGK--KDHSEDPN----- 700
DB 852 NNPKNNYTTISLVLT-----SHIND-----ILVYCKMKMMDNEVYALKY 892
QY 701 ---KNF-----KADEEVEETPAPEVYQVETEKVEAQLKEAEV-----LTA 739
DB 893 LGGNYILDLNDYTLRLDGINKVELISKQARELHEKMKDKKVPVLPITIGSNDLPLTL 952
QY 740 KYTDSLLKANATETTLAAGLRNNLTLLQIMDNNSINAAEKL-----LAILKGNSSSVSKEK 794
DB 953 EPLNSLYKADAGAVAA-----IPENTKVTYSKAKKISTKYKGLAKOHSSSSSVASGS 1004

QY 795 IN 796
DB 1005 VS 1006

RESULT 12
PTNL RAT
ID PTNL RAT STANDARD; PRT; 1175 AA.
AC 062728; 062732;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein tyrosine phosphatase, non-receptor type 21 (EC 3.1.3.48)
DE (Protein-tyrosine phosphatase 2E).
GN PTNP21 OR PTP2E.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2E).
RC STRAIN=Sprague-Dawley;
RX MEDLINE=95104449; PubMed=7805871;
RA L'Abbe D., Barville D., Tong Y., Slocco R., Masson S., Ma S.,
RA Fantus G., Shen S.H.,
RT Identification of a novel protein tyrosine phosphatase with sequence
RT homology to the cytoskeletal proteins of the band 4.1 family.";
RL FEBS Lett. 356:351-356 (1994).
CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC tyrosine + phosphate.
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q62728-1; Sequence=Displayed;
CC Name=2E;
CC IsoId=Q62728-2; Sequence=VSP_000498;
CC -1- TISSUE SPECIFICITY: PARTICULARLY ABUNDANTLY IN ADRENAL GLANDS.
CC -1- SIMILARITY: Contains 1 FERM domain.
CC -1- SIMILARITY: BELONGS TO THE NON-RECEPTOR CLASS OF THE PROTEIN-
CC TYROSINE PHOSPHATASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL: U17971; AAA62153.1; -
DR EMBL: U18293; AAA62154.1; -
DR PIR: S51005; S51005.
DR HSSP: Q06124; 2SHP.
DR InterPro: IPR000299; Band 4.1.
DR InterPro: IPR000387; TYR_phosphatase.
DR InterPro: IPR000242; Tyr_PP.
DR Pfam: PF00373; Band_41; 1.
DR Pfam: PF00102; Y_phosphatase; 1.
DR PRINTS: PR00700; PRTYPHTASE.
DR SMART: SM00295; B41; 1.
DR SMART: SM00194; PTPC; 1.
DR PROSITE: PS00660; FERM_1; 1.
DR PROSITE: PS00661; FERM_2; 1.
DR PROSITE: PS00657; FERM_3; 1.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE: PS00055; TYR_PHOSPHATASE_2; 1.
DR PROSITE: PS00056; TYR_PHOSPHATASE_2; 1.
KW structural protein; cytoskeleton; Hydrolase; Alternative splicing.
FT DOMAIN 23 308 FERM.
FT DOMAIN 922 1175 PROTEIN-TYROSINE PHOSPHATASE.
FT ACT_SITE 1109 1109 BY SIMILARITY.

```

FT VARSPLIC 1 839 Missing (in isoform 2E)
FT SEQUENCE 1175 AA; 133411 MW; 82A664F1C05EC7 CRC64;
SQ
Query Match 3.6%; Score 149.5; DB 1; Length 1175;
Best Local Similarity 19.4%; Pred. No. 1.5;
Matches 138; Conservative 87; Mismatches 273; Indels 213; Gaps 30;

OY KTEINTPEVSKREGINAEQIVIKITDOGYVTS-----HGHHYHNGKVPY----- 75
DB EAEMLYWQEVERMDYGEESYPAK-DSQSDISISACLDGIFVKKNGSPVFWEMDIA 260
OY 76 -----DAIIEBELKKDPNYKLEDEIVENGVYIKDGYVYVYLKDAADAVRTKE 130
DB 261 NMSHNKSEFALELANKEEIIQOTEDMET-----ACYVRLCVARHAKFRLNQC 309
OY 131 EINROKQSHQRECGTTPRNDGVALANSQ-----GYYTDDGYIFNASDII 177
DB 310 NLQTOAATLNSVRRGSSSR-----MSLPKQPYAMPPPDOLHYNGHYT-----EPF 355
OY 178 EDTGDAIYVPHGDHNYIPKNELSASELAAPISGKGNLSNRTY--RQNSDNTSRT 235
DB 356 ASSQDNVFPVKNNGVCHSQTSLDRTQID-----LSGR--IRNGSVSAHSTNSINTPQ 408
OY 236 MWVPS--VSNPQTNTNTSNNSTNSQASQ-----SNDIDLKQLYK-LPLSQRH 283
DB 409 YLQSPMSSNBPISGSDVMRPDYIPIHSHSALIPISYPTPTPYEESVWKRLNGWYHARRH 468
OY 284 VES-----DGLVFDPAQITSRTAGVAVPHGDHYH-----IPYSQ 319
DB 469 SHSLRNINIGSSYAYSRPDALVYSQPEI--REHPLHASPOSAAHYFPNLNYSFHSQAPY 526
OY 320 MSELERARIIPLRYRNNHWVPSRPE-----QSSPQT-----PESSPQ 362
DB 527 PVERRPVVAV-----SVPELTIVQLOADYPAENIMRTQVYRPPPYPR 573
OY 363 PA---PNLK-----IDSNSLVQVLRKVGEGYVEEKGISRYVPAKDLPSFTVNLBSK 414
DB 574 PANSTPDLNRHLIYSSSNPDILTRVYHNSVQ--TQEOBSLPVAHSLQGVSEPLTAARAH 631
OY 415 LSKQESV-----SHTLTAKENAVAR-----DOEFYKAVNLLTEAH- 451
DB 632 LQKRSNIEIAGLTHGFEGLRLKEETMSASADVADRFTSAGSSQSVFSDKVGQEGTEBOG 691
OY 452 KALFNKGRNSDFOL---DKLLERLNDESTNKEVLVDLL-----AFLAP 494
DB 692 SGGYSHKSLSDATMLHSSEDEDEDDSSREHNVSEPRLLTAAPSQEOQLNYPCASVTP 751
OY 495 ITHPERLGPNSQIYETDEVARIAQLADKYTSSDGYIFDEHDILSDEGDAVYTPHMGSH 554
DB 752 VTGPHLIEPKSHV--TEPEKAKDISPVHLVME-----THQPRRH 790
OY 555 WIGKDSLSDKEKVAQAAYTKEGILPSPDADVKANPTGSAALAYNEVGK 605
DB 791 GLTFPSMESDILTTSGRYRARDSL-----KKRPVSDLLSGKKNIVEG 833

RESULT 13
RBP1 PLAVB STANDARD; PRT; 2869 AA.
AC Q00798;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Reticulocyte binding protein 1 precursor.
GN RBP1.
OS Plasmodium vivax (strain Belm).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=31273;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=9231538; PubMed=1617731;
GA Galinski M.R., Medina C.C., Ingravallo P., Barnwell J.W.;

```

```

RT "A reticulocyte-binding protein complex of Plasmodium vivax
RT merotoiles."
RL Cell 69:1213-1226 (1992).
CC -1- FUNCTION: INVOLVED IN RETICULOCYTE ADHESION. SPECIFICALLY BINDS TO
CC HUMAN RETICULOCYTE CELLS.
CC -1- SUBUNIT: Homodimer (Potential).
CC -1- SUBCELLULAR LOCATION: Membrane-bound.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC
CC DR EMBL; M8097; AAA29743.1;
CC KW Malaria; Receptor; Signal; Transmembrane.
CC FT SIGNAL 1 17 POTENTIAL.
CC FT CHAIN 1 2869 RETICULOCYTE BINDING PROTEIN 1.
CC FT DOMAIN 18 2807 EXTRACELLULAR.
CC FT TRANSMEM 2808 2826 POTENTIAL.
CC FT DOMAIN 2827 2869 CYTOPLASMIC.
CC FT SITE 1030 1032 CELL ATTACHMENT SITE (POTENTIAL).
CC FT SITE 2599 2601 CELL ATTACHMENT SITE (POTENTIAL).
CC SQ SEQUENCE 2869 AA; 330213 MW; B9DBE442205EBCFF CRC64;

Query Match 3.6%; Score 149; DB 1; Length 2869;
Best Local Similarity 17.2%; Pred. No. 5.6;
Matches 166; Conservative 148; Mismatches 336; Indels 314; Gaps 42;

OY 24 KOATOKTENLTPDVSKEGINAEQ-----IVIKITD-----OGY--VTSH 62
DB 1702 KESTYKMLTYQNMESR---INVEGSLTIDKXITPIDNLLKMKQYEBGLLOKIKEN 1758
OY 63 GDHYHYNGKVPYAIIEBELMKDPN-----YKLEDEIVNEKGGVYIKVKGKYYV 116
DB 1759 AD-----KRSNFPVLSSEINALLDPSTSIPIKXLEWDMTGDK-NYGVKNMEIHGEF 1812
OY 117 LKD---AAHADN-----VRYKEINROKQESHQRE----- 144
DB 1813 TKSYNLIEHLNSNTDYSVTFPEKAQSLRELAKEBEHLRRREBAIFLNDIKKVESLKL 1872
OY 145 -----GGTPRNDGVALARSQGRITTD-----GYFNASDIIEDTG 181
DB 1873 LKEMMKVSAEYEGMKDHTSVSOLVODMKIYDELTLTNDISCSVANNVSVIVKVK 1932
OY 182 DAYIVPHGDHNYIPKNELSASELAAPISGKGNLSNRTYRQNSDNTS--RTMW 238
DB 1933 ES---KHADYR--DANSWYESWTLTANVFLSDAKISSGMEFPAEMKSNFKTDLLEIF 1987
OY 239 PVSNPCTNTNTSNNSN-----TNSQASQNSNDISLKLQLYKPLSGRYHESDGLV 290
DB 1988 SVIANSNELLKKEIQDSDNDVYQKERESEOLAKQVTDIYNYIK--LKNFEKKEEAKNKE 2045
OY 291 FDPQOITSRTARGVAVPHGDHYHFIYSQW---SELEERARIILPLRYRNNHWVPSRPE 347
DB 2046 EVVSEKVRALKRISQVEGIRCHENFHRLLDNTBEELNKKQVTTI-YR-----DKKSE 2098
OY 348 QPSQPTPEPSPGQOPANLKIDSNSLVQVLRKVGEGYVEEKGISRYVPAKDLPSFT 407
DB 2099 RES-----GLQENEN-EMNTYNSITQL-----EGIVSAGSEKEDIKLERSENE 2143
OY 408 VKNLESLSKQESVSHTLTAKKENAVARPDQEFYDKAVNLLTEAHKALFXNKGNSDPOAL 467
DB 2144 MRNISEKIS-----TIDSKVIEMNSTIDELYLKGNH--COAHWISLISYTAN---MKT 2191
OY 468 DKLLERLNDESTNKEKVLVDLLAFLAIPTHPERLGRKNSQIYEYEDVRIAQLADKYTTS 527
DB 2192 SKKLIMTKENKENTKCD-----YIKQNSST 2219
OY 528 DGYI-----DDEHDILSDEGDAVYTPHMGSHWIGKDSLSDKEKVAQAAYTKE 575

```

```

Db 2220 DGVEETLKPFYSGKLTFFSSASEIVQADTVSYNFAKHE-----KESLNARIDIKKELYLPH 2275
Qy 576 KGIILPSPDADYKANTG-DSAAIYNRVYKGR-----IPVRLPYMVEHVEV-- 624
Db 2276 Q-----NSDISIVGQVQNMALADKLANEGRNDELRYNRSERKLQOM-EHSIDVKK 2327
Qy 625 -----KNGNLIIP----- 632
Db 2328 PMIELHKGNNETNNKSLLEKEKKLKSVDNHHMSMEAEMLKNGLKYPESQVQNNIYSVI 2387
Qy 633 -----HCHY-----HNKPFAPDHT-----YAPNGYTLIEDLPAIT 665
Db 2388 EAEVKTLEEDRDYDGNVYIVIEEHKKQFSLIDRTALMDIDIEIFKKNENYLN----- 2440
Qy 666 KYVEHPDERPHS-NDGMCNASPHVLGKYDSEDPKPKFADPEPVETPAEP----- 717
Db 2441 ---MEVNTETIRVNDYIEIKTKVLQVQTEYQIILENKQNDMLQNIPLKVSIIIEYF 2497
Qy 718 EWPQVETEKVEAQLKAEVILAKVTD--SLKANATETLAGLRNNLTLOIMDNNSIMAAE 775
Db 2498 ENVKKKKESILNDLYEQERLL-KIGHLDEIKKNVTETLSSVYIDQKMEMSKN-LLEKK 2555
Qy 776 EKLL 779
Db 2556 SKKM 2559

RESULT 14
CUT7 SCHPO STANDARD; PRT; 1085 AA.
ID CUT7 SCHPO
AC P24339;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Kinesin-like protein cut7.
GN CUT7 OR SPAC25G10.07C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OC NCBI_TaxID=4896;
RN NCBI_TaxID=4896;
RP SEQUENCE FROM N.A.
RX MEDLINE=91015362; PubMed=2145514;
RN Hagan I., Yanagida M.;
RT "Novel potential mitotic motor protein encoded by the fission yeast
RT cut7+ gene."
RL Nature 347:563-566(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21848401; PubMed=11859360;
RN Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RN Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RN Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RN Collins M., Connor R., Cronin A., Davis P., Felwell T., Frazer A.,
RN Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RN Holtroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RN James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RN Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeh C.,
RN Oliver K., O'Neill S., Pearson D., Quail M.A., Rabbinowitsch E.,
RN Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RN Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RN Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RN Woodward J., Volkhardt G., Aert R., Robben J., Grynopre B.,
RN Weljens J., Volckaert G., Aert R., Robben J., Grynopre B.,
RN Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RN Borzym K., Langer I., Beck A., Leinhardt R., Reinhardt R., Pohl T.M.,
RN Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,
RN Goffeau A., Cadieu E., Dreano S., Gloux S., Leleure V., Mottier S.,
RN Gallbert F., Aves S.U., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RN Lucas M., Rochet M., Gallardin C., Tallada V.A., Garzon A., Thode G.,
RN Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,

```

```

RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Foreburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
CC -1- FUNCTION: COULD BE A SPINDLE POLE BODY MOTOR. ON TRANSITION FROM
CC G2 TO M PHASE OF THE CELL CYCLE, THE SPINDLE POLE BODY DUPLICATES;
CC THE DAUGHTER POLE BODIES SEED MICROTUBULES WHICH INTERDIGITATE TO
CC FORM A SHORT SPINDLE THAT ELONGATES TOWARD THE NUCLEUS AT
CC METAPHASE. MUTATIONS AT CUT7 BLOCK SPINDLE FORMATION.
CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY. BIMC
CC SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X57513; CAA40738.1; -.
CC EMBL: Z70691; CAA94636.1; -.
CC PIR: T38378; T38378.
CC HSSP: P17119; 3KAR.
CC GenDB SPombe: SPAC25G10.07C; -.
CC InterPro: IPRO01752; kinesin_motor.
CC Pfam: PF00225; kinesin; 1.
CC SMART: SM00380; KINESINHEAVY.
CC PRINTS: PR00129; KISC; 1.
CC PROSITE: PS00411; KINESIN MOTOR DOMAIN1; 1.
CC PROSITE: PS00067; KINESIN MOTOR DOMAIN2; 1.
CC Motor protein: Cell division; Microtubules; ATP-binding; Coiled coil;
CC Mitosis; Cell cycle; Phosphorylation; Repeat.
CC DOMAIN 70 435 KINESIN-MOTOR (BY SIMILARITY).
CC DOMAIN 436 604 COILED COIL (POTENTIAL).
CC DOMAIN 715 740 COILED COIL (POTENTIAL).
CC DOMAIN 897 955 COILED COIL (POTENTIAL).
CC NP_BIND 159 166 ATP (BY SIMILARITY).
CC REPEAT 967 998
CC REPEAT 999 1010
CC MOD_RES 1011 1011
CC CONFLICT 34 61 PHOSPHORYLATION (BY CDC2) (BY
CC SIMILARITY).
CC SASNPKRREPITNGVDRDNTSPT ->
CC LRAIIGNVSLILTL (IN REF. 1).
SQ SEQUENCE 1085 AA; 122133 MW; 5669277875559058 CRC64;
Query Match 3.5%; Score 145.5; DB 1; Length 1085;
Best Local Similarity 18.6%; Pred. No. 2.3;
Matches 182; Conservative 136; Mismatches 357; Indels 301; Gaps 41;
Qy 23 GKATQKTENTLTPDEVSREGINAE-----QIVIKI--TDQGYVSHGHYHYNN 70
Db 159 GQTGKGYTYMSGD-LSDSGDIGSAGILPRALYQLPSLDSNQEVAVK-CSYELVN 216
Qy 71 GAVPDAIISSELMKDNVYKIKED-----IVNEVKGQYVYKDGKYYVYKDAAHADN 125
Db 217 ERI-RDLVSEEL-RKPARVEEDTSRRGNVITGIESY-IKNAGDGLRLREGSHRQ 272
Qy 126 VATKEINROKQHS-----QREGCTPRNDGAVLAASQGRYTTDDGYTFNAPSIDET 180
Db 273 VAATKNDLSSSHSIFITTLARKVSSGMDTETNLTINN-----NSDDLIR-- 319
Qy 181 GDATVYPGHDHYVYPKNELSASELAAPLISGRGNLSNST----- 223
Db 320 -----ASKLHWVDLAGSENI--GRSGAENKARRETMINGSLTLTGRVI 361
Qy 224 -----YR-----RQNDNSTRTWVSVSNPGTNTNTGNSNNTNSQAQSS 264
Db 362 NALVEKAHHIPRESKLTLLLODSIGGKTKSMIVTVS---STNNLEETSTLEVAAYA 418
Qy 265 NDIDSLKQLYKLPISQHVESDGLVFPDAI-----TSRTARGAVPHGDHYHFIYEQ 319

```

```

Db 419 K---SIRRNQNNQLVFRKVLIKDLVIERLKNDLNTARRKNGVYLAESTYKEIMDRVQ 475
Oy 320 MSEL-----EERIARIIPLRYSNMHVPPSPROPSPQTPPSPROPAPPLKTI----- 369
Db 476 NKDLCOBARKLEVLDTLVKKS-----SRQLOQVSKSNQEHKKEVEALDLOLVNSSTE 529
Oy 370 -----DSNSIVSOLVRKGEVYVEEKGISRYVPAKPLPSTVANNLESKLSKQESVSH 424
Db 530 LESVSENEKLEKNEVLVEIEKKRKKEETBEAKITTYATLDSQVRESKEYIASLYEKLDRT 589
Oy 425 LTAKKENVAPRDQEFYKAVNL-----LNDSESTKEKIVDLDLAFIAIT 496
Db 590 ERNNKEN-----ENNFMNLFKNILTMLRSFHGSFTDETNGYFTLLNDPNSMEELLINTSHN 645
Oy 452 KALFRNKRNDFOALDKLER-----LNDSESTKEKIVDLDLAFIAIT 496
Db 646 QLLISMRTKITEHPSLDLALQSARSSCAVPNSSLDLIVSELKDSKNSLIDLALHSLQDIS 705
Oy 497 -HPERLGKPN-----QIEYTEDE-----VR--IAQL 520
Db 706 MSSQTLGNGISSELELOKDKESYRQVQLRSILYNLQHTHEESQKELMTGVRNDIDL 765
Oy 521 ADKYTTS--DGYIFDEHDIISDEGDAYVTPHMGSHWIGKDSLQKEXVAAQAYTKKGI 578
Db 766 VKCTTSLNDADII--LSDYISDQSKFESKQODLIANIGKIYVNLQONESLYTK--- 820
Oy 579 LPSPSDAVKANPTGDSAAAI--YRVVKGKRIPLVR-----LPMVETTVYVK 625
Db 821 -----ADILHSLNDTNSIRKANEIMNRRSEEFRLNNAQAEIVGANKERIQKTV-- 872
Oy 626 NGNLIIPKPDH--YHNIFKAMFPDDHTYKAPNG-----YTLEDLFATIKYVEHEDRP 676
Db 873 NSQSLIDSKSKAIHNSNSMTHCLALAESQKQVNLVQITDLRLQVK---E 923
Oy 677 HSNDSMGNASEHVLGKDHSEDPNKADEEVEETPAEPVQVETEKVYEAQLKEAV 736
Db 924 HSEDNTKEHQQLDLLESLVGNNDNL-----IDSIKTPHELIQ----- 962
Oy 737 LLAKTDSLSK-----ANATETLAGIRN--NLTLQIMDNNSIAEA-----EKLIA 780
Db 963 ---KITDVLVKGTSLAHNTTLLGIGDESLCNLETTIEDTSLVLETTGDTPSKRELPA 1019
Oy 781 LKGSNPSSVSKEKIN 796
Db 1020 TPSTWTRDSSLIKETTN 1035

RESULT 15
YNU1_YEAST STANDARD; PRT; 1240 AA.
AC P53935;
DT 01-OCT-1996 (rel. 34, Created)
DT 01-OCT-1996 (rel. 34, Last sequence update)
DT 01-OCT-1996 (rel. 34, Last annotation update)
DE Hypothetical 141.5 kDa protein in YP53-RHO2 intergenic region.
GN YNL091W OR N2231.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=96367601; PubMed=8771715;
RA Garcia-Camalejo J.M., Boskovic J., Jimenez A.;
RT "Sequence analysis of a 1.2 kb fragment of Saccharomyces cerevisiae
RT chromosome XIV that includes the YP53, TRN14u and gar m2 genes and
RT four new open reading frames."
RL Yeast 12:539-608(1996).
CC -I- SIMILARITY: TO S.POMBE SPAC2986.10C.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announce/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; X85811; CAAS9826.1; -
DR EMBL; 271367; CAAS95967.1; -
DR PIR; S52734; S52734.
DR SGD; S0005035; YNL091W.
DR GO; GO:0009651; P:salinity response; IMP.
DR K01001; P:protein.
FT DOMAIN 756
FT POLY-GU.
SQ SEQUENCE 1240 AA; 141513 MW; 3FE9D265822D5778 CRC64;

Query Match 3.5%; Score 145.5; DB 1; Length 1240;
Best Local Similarity 19.8%; Pred. No. 2.7; Indels 277; Gaps 35;
Matches 148; Conservative 85; Mismatches 236;

Oy 220 NSRTYRQNSDNTSRTNWVPSVSNPCTTNTNTSNTNSQASQSDNDISL-----KOLY 275
Db 4 NSKSKRRKN-----KSKQNNKKNQNSDPQSIPTQLVPRMEPELY 44
Oy 276 KLPLSQHVESDGLVDPDAQITRTARGVAVPHGDHYHFIPIYSOMSELERIARIIT---- 331
Db 45 -----HTESD--YPTSRVIRKAPNGDVIYE-----PINTDDKKERTANLTNNKD 87
Oy 332 -----PLRIR-SNHWPVDSRPREPSQPTPPEPSQGPAPNLKIDNSLSVSLVYK-- 382
Db 88 SMDASSIAFTLDSHW-----ESLSE-----EKKTLIRKEE--VEFVNIYNQ 130
Oy 383 -----VOEGVFEKGISR-----YVFAKDLPEST-----VNKLE-SKLSQ 418
Db 131 DDHSCSCVCGRRRLANDQEMERLYNTLYAMDKQDEFTPIKHLGIIKELQSKNOQ 190
Oy 419 ESVSHT-----LTAKKENVAPRDQEFYKAVN--LTFAHKALE 455
Db 191 NDISSTKGEVVKNFLSSSTVSLKEEVLHFQKQLSKQOAHNETADNTSLLEENLNINH 250
Oy 456 XNKR--NSDFOLDLRLINDESTNKEK-----LVDD 487
Db 251 INKTSSEISANFNVSVD--BELQOKYSNFTKFTISSHPKIAEEYVQMMYPNIRALTD 308
Oy 488 LL-----AFLAPITHPERLG--KPNQSEVTEDEVRIALQALD-----KYTT 526
Db 309 LANSNGGFLNAIEDPFRDQIQASKDDSI--TEDDASSITDLDPKEFTMLHSGKPLT 366
Oy 527 SDGYIFDEHDIISDEGDAYVT-----PHMGSHWIGKDSLQKEXVAAQAYTK 574
Db 367 EDEYADLQRMIAEMTNAYDTASKKFDVQOLEKELFTRFMSGRDKSPFELIIQSEFNK 426
Oy 575 EKGILPSPDA-----DYKANPTGDSAAAIY-----NRKGEKRI 609
Db 427 FDGELGVSVALATLSGCFSSQSKDTSIDTSDIYDEDEDEDDYDYSEVAEDSESEVEGI 486
Oy 610 PLVRLPYVVEHTEVVKQGNLIPKDHVYHNKIFAMPDHTYKAPNGYTLLEDLFAITIKYV 669
Db 487 EAVKRP--EHD-EKSGIKRETLLSY-----DHHKRN----- 517
Oy 670 EHPDERPHSNDGMGNASEHVLGKDH-----SDPNKNFKADEEVEETPAEPVPOV- 722
Db 518 -HPHHVHYST--STHSEDELSSEBEYISDIELPHDPKHFHRRDDIIDGDEDEDEEDEN 573
Oy 723 -----ETEKVEAQLKEAVLAKYTDSSISKANATTTLAGLRNLTLMQMDN 768
Db 574 EGDDEEDPYDGLDETRLEGRKLIQIATKLIQSRIMASYHKAQ-----DN 622
Oy 769 NSIMAEKLIALLKGSNPSSVSKEK 794
Db 623 NRL-----KLLQLELBEKKRKREKEE 643

RESULT 16

```

JIP3_MOUSE
 ID JIP3_MOUSE STANDARD; PRT; 1337 AA.
 AC Q9ESN9; Q9EKU7; Q9E0B8; Q9ESN8; Q9ESP0; Q9ULH2; Q9ULH3;
 AC Q9R0U7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE C-jun-amino-terminal kinase interacting protein 3 (JNK-interacting
 protein 3) (JIP-3) (JNK MAP kinase scaffold protein 3) (Mitogen-
 activated protein kinase 8-interacting protein 3) (JNK/SAPK-associated
 protein 1) (USAP1) (Sunday driver 2).
 GN MAPK1P3 OR USAP1 OR JIP3 OR SYD2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM 1A), FUNCTION: PHOSPHORYLATION, AND
 RP INTERACTION WITH MAPK8; MAPK9; MAPK10; MAP2K4 AND MAP3K1.
 RC TISSUE=Brain;
 RX MEDLINE=99455010; PubMed=10523642;
 RA Ito M., Yoshioaka K., Akechi M., Yamashita S., Takamatsu N.,
 RA Sugiyama K., Hibi M., Nakabeppu Y., Shiba T., Yamamoto K.-I.,
 RT "USAP1, a novel jun N-terminal protein kinase (JNK)-binding protein
 that functions as a scaffold factor in the JNK signaling pathway.",
 RL Mol. Cell. Biol. 19:7539-7548(1999).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 1A, 1B, 1C, 1D AND 1E), AND TISSUE
 RP SPECIFICITY.
 RC TISSUE=Brain;
 RX MEDLINE=20480689; PubMed=11024282;
 RA Ito M., Akechi M., Hirose R., Ichimura M., Takamatsu N., Xu P.,
 RA Nakabeppu Y., Tadayoshi S., Yamamoto K.-I., Yoshioaka K.;
 RT "isoforms of USAP1 scaffold protein generated through alternative
 RT splicing.",
 RL Gene 255:229-234(2000).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORMS 1C AND 3A), FUNCTION: INDUCTION, TISSUE
 RP SPECIFICITY, SUBCELLULAR LOCATION, PHOSPHORYLATION, MUTAGENESIS OF
 RP ARG-205, PRO-206, THR-207, SER-208, LEU-209, THR-266, THR-276 AND
 RP THR-287, AND INTERACTION WITH MAPK8IP2; MAPK8; MAPK9; MAPK10; MAP2K7
 RP AND MAP3K1.
 RX STRAIN=C57BL/6; TISSUE=Brain; and Healt;
 RC MEDLINE=20094982; PubMed=10629060;
 RA Kelker N., Gupta S., Dickens M., Davis R.J.;
 RT "interaction of a mitogen-activated protein kinase signaling module
 RT with the neuronal protein JIP3.",
 RL Mol. Cell. Biol. 20:1030-1043(2000).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 1C), SUBCELLULAR LOCATION, AND INTERACTION
 RP WITH KLC1
 RC STRAIN=C57BL/6; TISSUE=Brain;
 RX MEDLINE=20560733; PubMed=11106729;
 RA Bowman A.B., Kamal A., Ritchings B.W., Philp A.V., McGarail M.,
 RA Gindhart J.G., Goldstein L.S.B.;
 RT "Kinesin-dependent axonal transport is mediated by the Sunday Driver
 RT (SYD) protein".
 RL Cell 103:583-594(2000).
 RN [5]
 RP SEQUENCE OF 1240-1337 FROM N.A.
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strusberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E.,
 RA Brownstein M.J., Ustin T.B., Toshynki S., Carninci P., Pange C.,
 RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Millany S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.T., Skalska U., Smalins D.E.,
 RA Schermer A., Schein J.E., Jones S.J.M., Maitra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.",
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [6]
 RP INTERACTION WITH KLC.
 RC TISSUE=Brain;
 RX MEDLINE=21135887; PubMed=11238452;
 RA Verhey K.J., Meyer D., Deehan R., Blenis J., Schnapp B.J.,
 RA Rapoport T.A., Margolis B.;
 RT "Cargo of kinesin identified as JIP scaffolding proteins and
 RT associated signaling molecules.",
 RL J. Cell Biol. 152:959-970(2001).
 CC -1- FUNCTION: The JNK-interacting protein (JIP) group of scaffold
 CC proteins selectively mediates JNK-signaling by aggregating
 CC specific components of the MAPK cascade to form a functional JNK
 CC signaling module. May function as a regulator of vesicle
 CC transport, through interactions with the JNK-signaling components
 CC and motor proteins.
 CC -1- SUBUNIT: Forms homo- or heterooligomeric complexes. The central
 CC region of Mapk8ip3 interacts with the C-terminal of Mapk8ip2 but
 CC not Mapk8ip1. Binds specific components of the JNK signaling
 CC pathway namely Mapk8, Mapk9 and Mapk10 to the N-terminal region,
 CC Map2K4 and Map2K7 to the central region and Map3K1 to the C-
 CC terminal region. Binds the TPR motif-containing C-terminal of
 CC kinesin light chain, pre-assembled Mapk8ip1 scaffolding complexes
 CC are then transported as a cargo of kinesin, to the required
 CC subcellular location.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic, localised in the soma and
 CC growth cones of differentiated neurites and the Golgi and vesicles
 CC of the early secretory compartment of epithelial cells.
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event-Alternative splicing; Named isoforms=6;
 CC Name=1c; Synonyms=3b;
 CC IsoId=Q9ESN9-1; Sequence=Displayed;
 CC Name=1a;
 CC IsoId=Q9ESN9-2; Sequence=VSP_002775, VSP_002777;
 CC Name=1b;
 CC IsoId=Q9ESN9-3; Sequence=VSP_002776, VSP_002777;
 CC Name=1d;
 CC IsoId=Q9ESN9-4; Sequence=VSP_002775;
 CC Name=3a;
 CC IsoId=Q9ESN9-5; Sequence=VSP_002778, VSP_002779;
 CC Name=1e;
 CC IsoId=Q9ESN9-6; Sequence=VSP_002776;
 CC -1- TISSUE SPECIFICITY: Highly expressed throughout many regions of
 CC the brain and at lower levels in the heart, liver, lung, testes
 CC and kidney. All isoforms have been identified in the brain,
 CC Mapk8ip3A is also expressed in the spleen and lung.
 CC -1- INDUCTION: Expressed in neurites 5 days following initiation of
 CC nerve growth factor Ngf induced differentiation. Ngf withdrawal
 CC results in the down-regulation of Mapk8ip3 protein by caspase-
 CC mediated cleavage.
 CC -1- SIMILARITY: BELONGS TO THE JIP SCAFFOLD PROTEINS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL: AB005662; BAA85874.1; -
 CC EMBL: AB043124; BAB16675.1; -
 CC EMBL: AB043125; BAB16676.1; -
 CC EMBL: AB043123; BAB16674.1; -
 CC EMBL: AB043129; BAB16685.1; -
 CC EMBL: AB043126; BAB16685.1; JOINED.

```

DR EMBL: AB043127; BAB16685.1; JOINED.
DR EMBL: AB043128; BAB16685.1; JOINED.
DR EMBL: AF178637; AAF26843.1; -.
DR EMBL: AF178636; AAF26842.1; -.
DR EMBL: AF262046; AAG36931.1; ALT_INIT.
DR EMBL: BC004003; AAH04003.1; -.
DR MGD: MGI:1353598; MAPK8IP3.
DR GO: GO:0005737; C:cytoplasm; IEP.
DR GO: GO:0005078; F:kinase binding activity; IPI.
DR GO: GO:0005078; F:MAP-kinase binding activity; IPI.
DR GO: GO:0019901; F:protein kinase binding activity; IPI.
DR GO: GO:0046328; P:regulation of JNK cascade; IDA.
DR GO: GO:0016192; P:vesicle-mediated transport; IDA.
KW Alternative splicing; Phosphorylation; Coiled coil.
FT DOMAIN 58 177
FT MOD_RES 437 555 COILED COIL (POTENTIAL).
FT MOD_RES 266 266 PHOSPHORYLATION (BY MAPK).
FT MOD_RES 276 276 PHOSPHORYLATION (BY MAPK).
FT MOD_RES 287 287 PHOSPHORYLATION (BY MAPK).
FT VARSPLIC 201 201 Missing (in isoform 1a and isoform 1d).
FT VARSPLIC 201 201 /FTId=VSP_002775.
FT VARSPLIC 201 201 S-> SPROSMKRS (in isoform 1b and isoform 1e).
FT VARSPLIC 219 249 /FTId=VSP_002776.
FT VARSPLIC 410 415 Missing (in isoform 1a and isoform 1b).
FT VARSPLIC 505 513 Missing (in isoform 3a).
FT MUTAGEN 205 205 /FTId=VSP_002778.
FT MUTAGEN 205 205 Missing (in isoform 3a).
FT MUTAGEN 206 206 R->G: RESULTS IN INHIBITION OF JNK BINDING.
FT MUTAGEN 207 207 P->G: RESULTS IN INHIBITION OF JNK BINDING.
FT MUTAGEN 208 208 T->G: RESULTS IN INHIBITION OF JNK BINDING.
FT MUTAGEN 209 209 S->G: RESULTS IN INHIBITION OF JNK BINDING.
FT MUTAGEN 266 266 L->G: RESULTS IN INHIBITION OF JNK BINDING.
FT MUTAGEN 276 276 T->A: RESULTS IN LOSS OF PHOSPHORYLATION OF MAPK8IP3; WHEN ASSOCIATED WITH A-276 AND 287. DOES NOT EFFECT BINDING OF COMPONENTS OF THE JNK PATHWAY.
FT MUTAGEN 287 287 T->A: RESULTS IN LOSS OF PHOSPHORYLATION OF MAPK8IP3; WHEN ASSOCIATED WITH A-286 AND A-287. DOES NOT EFFECT BINDING OF COMPONENTS OF THE JNK PATHWAY.
FT MUTAGEN 312 312 T->A: RESULTS IN LOSS OF PHOSPHORYLATION OF MAPK8IP3; WHEN ASSOCIATED WITH A-266 AND A-287. DOES NOT EFFECT BINDING OF COMPONENTS OF THE JNK PATHWAY.
FT CONFLICT 312 312 K->R (IN REF. 4).
FT CONFLICT 376 376 F->L (IN REF. 3; AAF26843).
FT CONFLICT 561 561 E->K (IN REF. 4).

Query Match 3.5%; Score 144; DB 1; Length 1337;
Best Local Similarity 19.4%; Pred. No. 3.6; Indels 271; Gaps 40;
Matches 164; Conservative 112; Mismatches 298;

81 EELLKKDKNYK-LKDEDIVNEVGKGVYIKVDGKYVY-----LKDAAHADNV-RTKEEI 132
97 EKALAKQAEKFEIREDALQEKELQLOVE--HIEFQRLQELAKYADIDSLREBE 154
133 NROKOE-----HSQRE-----GGTPRNDGAVALARSGRATTDDGYIF 171
155 SEMKKEYNALHQRHTEMIOQYVEHIERSKMQGVGSGGQESSLP-GRSKKEPTSLANEP 213
172 NASDIIEDTGDAYIVPHGDHYIYIKNELSASELAAAEFLSGRGNLSRTRYRQNSDN 231
214 LADGVVRAQMGKGLVPAGDHHM-----LSDILOQSSSSSYQCPN-DE 254
232 TSRTWVSVSNPGTNTNTSNTNSNTNSQASQNDIDSLKLQLYLT----- 277

```

```

DB 255 MSESQSSAAATPSTGT-----KENTPTSSVPSAAVTPLENSIQPLGDEVSTKNNKQAR 310
OY 278 -PLSQRHVESGLVFPDAQITSRARGVAVPHGHHYIFIPISQMSLEBERARIIPRYR 336
DB 311 EKRSRNNE-----VQVT-QEMRVNVSIGMGS-----DEWSDVDIISTPELDV- 354
OY 337 SNHWPPSRPQSPQPP-----BSPSP----- 361
DB 355 -----CPETRLERTSSPTQGVNKAFGINTDSLHYELSTAGSEVIGVDDEADLIGERSV 410
OY 362 -----QPAVNLKIDNSSLV-----SOLVRKV-----GEGYVEEKG-ISRNV 398
DB 411 RDPFGMGKEVGNLLE-NSQLLETQALNVKNDLAKVQJLGEQEVLT--KGELEAK 467
OY 399 FAKLPSETVNLSEKSKQESVSHTLTAKKENAVPDDQETDQAVNULTTAHAALEPKNK 458
DB 468 QAKVLENRIKELBEELRVK--SEAVTARE--PR-EEVEDVSYLTCLDKI PMAQR 521
OY 459 GRNSDFQALDKLRLNDESTNKEKLVDDLLAFAPITHPRLGKPNQIERYDEVARIA 518
DB 522 RRTFVEMARLME-----NOYK-----ERLWELQEAARWTE-MIRAS 559
OY 519 QIADKYTTSDDYITDEHDIIISDEGDAYVTPHMGSHWIGKQSLDKEKVAQAAYTK---- 574
DB 560 R-----EHPSVQEKKSTIMQFPRLF-----SSSSPPPAKRSVPSVNIH 600
OY 575 -----EKGILPRS--PDADVANKPTGDSAAATYINRYKKEKRIPLV 612
DB 601 YKSPPTAAGFSQRSHALCOISAGSRPLEFFDDCTSSARREQREQYRVR----- 652
OY 613 RLPVWEHTVAVKGNLIIIPKHQYHNIIKFAWPDQHTYK--APNGYTLDEL--FATIKY 667
DB 653 -----EH--VRN-----DGRLOACGMSLPATYKQSLSPNG-QQEDIRMGKNVPPV 694
OY 668 YVEHPDERPHSNDGWSNASEHVLGKDHSEDPNKNFK-----ADEEPVEETP 714
DB 695 YCRFLVEKDPSTKLKCAAGVLSGWMKPHEDSSNGPKVPGRDPLTCDEGESEPKSTYHP 754
OY 715 AEPEVQVETKEVNAQLKEAVLAKYTDSS--LKNATTTLAGLNNLTLQMDNNSI 771
DB 755 SPEKKAKKEPPEADATSSRWILSTLTSTSKVLIIDANQPTIVDQFTVCNAHVCISII 814
OY 772 MAEAE 776
DB 815 PAASD 819

RESULT 17
ID SYA MYCGE STANDARD; PRT; 900 AA.
AC P47534;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Alanyl-tRNA synthetase (EC 6.1.1.7) (Alanine--tRNA ligase) (Alars).
GN ALAS OR MG292.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmales; Mycoplasma.
OX NCBI_Taxid=2097;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Frieser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA Fleischmann R.D., Bult C.J., Kellavage A.R., Sutton G., Kelley J.M.,
RA Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA Nguyen D.T., Uterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA Tomb J.-F., Dougherty B.A., Bot K.F., Hu P.-C., Lucier J.S.,
RA Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.,
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
-!- CATALYTIC ACTIVITY: ATP + L-alanine + tRNA(Ala) = AMP +

```

diphosphate + L-alanyl-cRNA (Ala).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U39709; AAC71513.1; -
 CC PIR: C64232; C64232.
 CC TIGR: MG292; -
 CC HAMAP: MF_00036; -; 1.
 DR InterPro: IPR002318; tRNA-synt_2c.
 DR InterPro: IPR006193; tRNA-synt_Ala.
 DR Pfam: PF01411; tRNA-synt_2c; 1.
 DR PRINTS: PR00980; TRNASYNTHALA.
 DR TIGRFAMs: TIGR00344; alas; 1.
 DR PROSITE: PSS00860; AA_TRNA_LIGASE_I; 1.
 DR AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KM Complete proteome.
 SQ SEQUENCE 900 AA; 104301 MW; AA54520BFB3949A2 CRC64;
 Query Match 3.4%; Score 143.5; DB 1; Length 900;
 Best Local Similarity 19.7%; Pred. No. 2.2; Indels 159; Gaps 28;
 Matches 127; Conservative 109; Mismatches 250;
 273 QLYKPLSGRSHVESDGLVDPAPQITSRTARG---VAVPHGDHYHPIFYQSMSELEERIA 328
 DB 199 ELMWNVFSGFNNDGNGVNTLAKNDIDTAGIERLVSVLQNS---PINFDTDFLKLKI 253
 QY 329 RII---PIRYSNHNWVDSRPDPSPDPPTPP-----SPGQAPAPL 367
 DB 254 KIIKAFCEPKYDPNSYFPP---DPQKVKESQSYRIIADHPKAITFTTISGVLPGPN- 306
 QY 368 KIDSNSLSLQSLVR-----KVGEVYFEKGISYVFAKO-----LPSETYK----- 409
 DB 307 ---ERNYVVRRLRLRALACKLQNLMAITEKIIDEIINASYNYOHLAKAKETVAVVL 363
 QY 410 ---NLESKLSQESVSHLTAKKENVAPRDOEFYDKAVMLTEAHKALFXNKGNSDFOA 466
 DB 364 KEINAFNKITIDGLVLFKSVANNITLTQTLFQNLNETTFEPPEIIRLVNQGKLTIDWTV 423
 QY 467 LDKL-----LERLDESTNKEKLVDDLLAFLAIPTHPRRLCKPNSOI---EYTEDSVRIA 518
 DB 424 FQQLMAKHSRISKNNQITNFEKQINILVNFKTKSTFFYKKNKINAKVIGLFDENVLPVK 483
 QY 519 QLAADK-----YTTSDGVIPEHDI---SDEGDAVYV-----PHNGHSHW--I 556
 DB 484 EINNNSGYVFPDQTVLYATSGGQRDEGSCINHSNNNDKISFQGVFKPKNQHPHYFLV 543
 QY 557 GKDSLSDK-----EKVAQAAYTK---KGILPPSPDAVYK---ANPTGDSAAIYN 601
 DB 544 GSFKANDQVTLSDHDEWRKLAANNSEHLILALALOKELDPLIKSGAKSKAKATIDN 603
 QY 602 RVKGEKRIPLVRLPYWVEHTVEVKG-----NLIIPKHGY--HNI-KFAM 644
 DB 604 INRHILTRNELEKVENKIRSLIKOKISSKEIFTDPEGSQKLMIAVEEESQHEILRVIR 663
 QY 645 FDDHYK-----AENGTYLEDLPATIKYVHPDERPHSNNGWGNASHVLGKDXHSD 698
 DB 664 FGDYVEVEICGGTHVANATASIEDCFITDFYSL-----GAGWRILEIIS--NBT 709
 QY 699 PNKNEKADPEVEETPADE--VPOVETKEVAOLKEAE-----VLAKTDSLSKANA 750
 DB 710 INNYLKAENOKILQKSELEKVLSDISIFKVELKEQORLDPKLPKTIQ--LRDA 766
 QY 751 TETLAGLRNNLTQLIMDNNSMAEA-----EKLALLKSGNSPSV 790
 DB 767 SDTLALAKNDINQLKTKNYKVSQALALATSIKKQLSLVDENKSYV 811

RESULT 18
 ZIPI_YEAST
 ID ZIPI_YEAST STANDARD; PRT; 875 AA.
 AC P31111;
 DT 01-JUL-1993 (Rel. 26, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE Synaptonemal complex protein ZIPI.
 GN ZIPI OR YDR285W OR D9819.9.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BR1824-3B; PubMed=7916652;
 RX MEDLINE=9316142; PubMed=7916652;
 RA Sym M., Engelbrecht J.A., Roeder G.S.;
 RT "ZIPI is a synaptonemal complex protein required for meiotic
 RT chromosome synapsis."
 RL Cell 72:365-378 (1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,
 RA Favello A., Fulton L., Gattung S., Greco T., Kirsten U.,
 RA Kuchba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,
 RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,
 RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,
 RA Rifken L., Riles L., Taich A., Trevaekis E., Vignati D.,
 RA Wilcox L., Wohldman P., Vaudin M., Wilson R., Waterston R.;
 RL Submitted (Mar-1996) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR MEIOTIC CHROMOSOME SYNAPSIS AND CELL CYCLE
 CC PROGRESSION. MAY ACT AS A MOLECULAR ZIPPER TO BRING HOMOLOGOUS
 CC CHROMOSOMES IN CLOSE APPPOSITION. ZIPI MAY ENCODE THE TRANSVERSE
 CC FILAMENTS OF THE SYNAPTONEMAL COMPLEX.
 CC -1- SUBCELLULAR LOCATION: SYNAPSED MEIOTIC CHROMOSOMES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L06487; AAA35239.1; -
 CC EMBL: U51031; AAB64474.1; -
 CC PIR: S70115; S70115.
 CC SGD: S0002693; ZIPI.
 CC DR GO: GO:0005716; C:synaptonemal complex; IDA.
 CC DR GO: GO:0007126; P:meiosis; IMP.
 CC DR GO: GO:0007129; P:synapsis; IMP.
 CC KW Nuclear protein; Meiosis; Coiled coil.
 FT DOMAIN 177 333 COILED COIL (POTENTIAL).
 FT DOMAIN 397 438 COILED COIL (POTENTIAL).
 FT DOMAIN 456 752 COILED COIL (POTENTIAL).
 FT CONFLICT 55 55 T -> A (IN REF. 1).
 SQ SEQUENCE 875 AA; 100035 MW; 674F12625CD9DDFD CRC64;
 Query Match 3.4%; Score 142.5; DB 1; Length 875;
 Best Local Similarity 18.9%; Pred. No. 2.4;
 Matches 176; Conservative 117; Mismatches 304; Indels 333; Gaps 41;
 3 ELGLVQART-VKENNRVSYIDGKQATQKTEMLTPPEVSKRREGSINAQIYKITDGYVTS 61
 DB 86 ELGSPKTTSTIDQYKRII---KNDVAALENDTDEDFETTEVREVSQVAKETKE---S 137
 QY 62 HGD-----HYHYNGKRVYPDAL-----SEELLM----- 85
 DB 138 HGDPDNDEFTLLKDSKMHETVMTNGKAPLHTSINNSTGSDNVLLAEFTVTRIGCSNLKOE 197

```

QY 86 ----KDPYKXK-----DEDIVNEVKGVIKVDGKYVYVLD-----AAHADN 125
D 198 LOKOQODAKKVKRLOSAYASNDKINE-----KV-GKYSCELTLOERLATLSHKON 249
QY 126 VRTKEIRKQKQKSHQREBGGTPRNDGAVALARSGGRYTTDDGYFNASDIETDG-----181
D 250 QETKXKLDJRQNHQLYORRISG-----FKTS---IENLNKTIINDGKMK 290
QY 182 --DAIVPHGHYHPIPKNELSASELAFAFLSGRNLNSRTRBRNSD-----NTSR 234
D 291 EADDELKMKKKEIYVL-KRELDGCS-----GQISEKTI---KNSSLIQMGKRN 335
QY 235 TNNVPSVS-----NPGTTNTNTSN 254
D 336 BEMIKSINFSEDKAHHLLOFNKFEERVDLFEKKLOKHPVADKTLNGLRNTVELS 395
QY 255 SNT-----NSQASQSDNDISLLKQYKPLPSQRH-----VESDGLV-- 290
D 396 SNTETMLKQYEDIKENLEQKMSKDEMAKTINELSTQKGLIMGVOEELLTSSGNIQT 455
QY 291 ----PDPQITSTARGVA-VPHGDHYHPIPSQMSLEERLARIIPLYR 336
D 456 ALVSEMNTRQELLDDA---SQTAKNYASLENLVKAAYAETVQSNMEYEBRIKHLESER-- 510
QY 337 SNHWVPDRPQPSQPTPEPSPGPQAPANLKDSSSLVSQLVAKVGEVGFEEKGISR 396
D 511 -----STLSSQKNOIISLSG-----TK 527
QY 397 YVFAKDLPEFYKNLESK---LSKQESVSHLTAKKENAVPRDQFYD--KAYNLLTEAH 451
D 528 BAQYEDL---VKLEAKNIEISQISGKEOSLTKENENLNLKVKQDQLEKLANLNTT 583
QY 452 KALFENK---GRNSDQAL---DKLERLNDSTKKEKLVLDLAFAPITPHEPLKPN 505
D 584 KSNENKSSQSEIYKALVSENDTLKORIQOIVEIKENQKQD-----HTTKLEFQKN 637
QY 506 SQIETEDEVRIQALADKYTTSDGYIFPEHDIISDEGAYVYTPHMGSHMIGKDLSPKE 565
D 638 EQLQ--KLVNEVQVQ-----KAHELELEQNRHLKCEKKEKTEGVEELSIVK 683
QY 566 KVAAQAYTEKGIKLEPPSPDADYKA-----NPTGSAALVYRVKGEKRIPLVRLPY 616
D 684 TL-----KQOIVYKSEKQDITAEKLEQDLJESLEETKYLQOKVOSQK---ELEQ 733
QY 617 MVEHTVEKNGNLLIIPKHQHNINIKFAMPDDHTYKAPNGYLTEDLFAPIKYVHEHDERP 676
D 734 KIKLEBETKHNKRNEPSKKTQON---FTKPSDSPKXNATTSNLPNNNSAALHSPMKKC 788
QY 677 HSNQGMGNASEHVLGKDKHSEDPNNKF-----KADEPVAEETPAEPEVQVETEKYEA 729
D 789 PKVD-----HISKRINSSEKTSKFPNDEPDLSSSNDLLETNPSF---IQIKPRG 837
QY 730 QLEKA-----EVLAKVTSSLK 747
D 838 KIKKGSNCKMPPISRRKLLLEVEDQSLK 867

```

```

RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.
RX MEDLINE=97041712; PubMed=8886983;
RA Toda T., Niwa H., Nemoto T., Dhut S., Eddison M., Matsusaka T.,
RA Yanagida M., Hirata D.;
RT "The fission yeast stg+ gene is required for maintenance of growth
RT polarity and functionally interacts with protein kinase C and an
RT osmosensing MAP-kinase pathway.";
RL J. Cell Sci. 109:2331-2342(1996).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RX MEDLINE=21846401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,
RA Gentile S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.D., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odeil C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tvey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Voickaert G., Aert R., Robben J., Grymoprez B.,
RA Welljens I., Vanstrele E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lebrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambut R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leleau V., Motier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Bento J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Pallesen I., Potashkin J.,
RA Shpakovski G.V., Usery D., Barrell B.G., Nure P.;
RT "The genome sequence of Schizosaccharomyces pombe.";
RL Nature 415:871-880(2002).
CC -!- FUNCTION: REQUIRED FOR THE MAINTENANCE OF CELL SHAPE DURING
CC INTERPHASE. REQUIRED FOR LOCALIZATION OF CORTICAL ACTIN TO THE
CC GROWING TIPS BEFORE MITOSIS.
CC -!- SUBUNIT: INTERACTS WITH SERINE/THREONINE PHOSPHATASE PPE1, PROTEIN
CC KINASE C AND AN OSMOSENSING MAP KINASE.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- SIMILARITY: Belongs to the ribonuclease II (RNB) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announce/
CC or send an email to license@ebi.ac.uk).
CC
DR EMBL; D58421; BA2619.1; -
DR EMBL; AL031535; CA20748.1; -
DR PIR; T41099; T41099.
DR PIR; T45283; T45283.
DR GenDB_Spombe; SPCC16C4.09; -
DR InterPro; IPR001900; Ribonuclease_II.
DR Pfam; PF00773; RNB; 1.
DR PROSITE; PS01175; RIBONUCLEASE_II; FALSE_NEG.
KW Hydrolase; Nuclease; Exonuclease.
FT CONFLICT 5 5 F -> C (IN REF. 1).
FT CONFLICT 794 794 V -> G (IN REF. 1).
SQ SEQUENCE 1066 AA; 117602 MW; 07718780DB330C15 CRC64;

```

Query Match 3.4%; Score 142.5; DB 1; Length 1066;
Best Local Similarity 21.9%; Pred. No. 3.2;
Matches 110; Conservative 69; Mismatches 192; Indels 131; Gaps 25;

```

QY 186 VPHGDHYHPIPKNELSASELAFAFLSGR-----GNSNSRTYRQNSDNTSR 234

```

```

Db 55 LPLQOHMQLRHTGL---LPAVSSFVHGHRSSASAGVGMGNSQATI---PSNSPAV 107
Qy 235 TWVPSV-----SNPGTTNTNTSNNSTNSQASQNDIDSLKQLYKPLS----- 280
Db 108 SNMQPPTGGGQPLYPINFTTSVSASDSDFRNSPTVPS-----KFSINPSVATS 156
Qy 281 -----ORHVESDGLVDPQAQTSRTARGVAVPHGDYHPIPSQMSLE--ERIARIPL 333
Db 157 TNISPRRAKSHSV---ASVSSPNSHN-AVPTPAFVFPVNNASPLPALNTLPQLLRP 211
Qy 334 RFRSNHWVDSRPEQSPQPTPEPS-PC-----PPAPNLIKITSNSGLVQLYR 381
Db 212 RULDAQWRPSSLSQNTSPTHAANPSFPGITVTHTNSFRPEGGHRHRSSTGSL---SVG 268
Qy 382 KYGEGVFPEKGISR-YVFAKDLPESTVKULESKLSKQSVSHTLTAKKENVAPRQDEY 440
Db 269 SSGSGSSGSGSGSNPRKNTLSPYLPQSSIAL--LBERLVTGILVSKKN--RSDAV 322
Qy 441 DKAYNLLTEAHKALFYNKGRN---SDFOALDKL-----LBERLNDSTNKEK----- 483
Db 323 SVDG---LDAEYFICGSKDRNRALLEGDVVAIELLDVDEVWAGKLEKENRRRKP1STRG 379
Qy 484 LVDDLAFIAPITPHERLKPNSQIETDEVRIAQLADKYTSDGYTDEHDIIISDEG 543
Db 380 SFDNLRIDAVPPEVPR---SAIARDEQVEGQTL-----FLDQKQGLGDEXP 426
Qy 544 AYVTPHMGSHWIGKSLSDKEKVAQAQATYKKGILPSPDADVK-----ANPTG 593
Db 427 KYA---GH-----VAVLQAPQGVFSGITILPSSAANKKQTSNGGSSNSNG 475
Qy 594 DSAAIYNRVKGEKRIPLVRLP 615
Db 476 NDKPKLVMEKPSDKRVPLIAP 497

RESULT 20
D7_DICDI STANDARD; PRT; 850 AA.
ID D7 DICDI
AC P54682;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE CAMP-inducible prespore protein D7 precursor.
GN D7.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_TaxId=44689;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=95080502; PubMed=7988791;
RA Agarwal A., Sloger M.S., Oyama M., Blumberg D.D.;
RT "Analysis of a novel cyclic Amp inducible prespore gene in
RT Dictyostelium discoideum: evidence for different patterns of CAMP
RT regulation.";
RL Differentiation 57:151-162(1994).
CC -1- DEVELOPMENTAL STAGE: EXPRESSED SPECIFICALLY IN THE PRESPORE CELLS.
CC -1- INDUCTION: BY CAMP.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U25143; AAA73514.1.
DR DictyDb; DD02038.
KW Sporulation; Signal.
FT SIGNAL 1 24
FT CHAIN 25 850 POTENTIAL.
FT DOMAIN 470 475 CAMP-INDUCIBLE PRESPORE PROTEIN D7.
FT POLY-GLN.

```

```

FT DOMAIN 555 568 POLY-ASN.
FT CHAIN 728 738 POLY-GLN.
SQ SEQUENCE 850 AA; 95343 MW; 13BA634CCE7AA502 CRC64;

Query Match 3.4%; Score 142; DB 1; Length 850;
Best Local Similarity 16.7%; Pred. No. 2.5;
Matches 140; Conservative 122; Mismatches 296; Indels 280; Gaps 29;

Qy 15 NNRVSYIDGKATOK-TEML--TPDEVSKREGINAEQIV--IKITDQGVTHGHGHHYH 69
Db 24 SNQISEDVGAIAINQKLENIEKEVEDVVOFENV-ANQVIEELKEHQ-----ROEL 73
Qy 70 NGKVPYDAIISEE-----LTKDPNYKDKDIDVINEVGGYVIKVGXYVYLKPAADA 123
Db 74 LGEVNHDSLRDATYIHWLLDKIQSYLPRKDNKKYSKVEAASSGQNNIGSSIGSTGA 133
Qy 124 DWRTKEEIN--ROKQESHQREGGTPRNDGVALARSGRYTTDDGYIFNADIIEDTG 181
Db 134 STSPQFQSIGNGSGASQSGSGSTGGTGDSDSK-----TTNEAIIFFSKVSTTDQ 183
Qy 182 DAYIVPHGDHYIYIPKNELASBELAAEAFSLSGRNLNSRTYRRQNSDNTSRTNVP 241
Db 184 ESI-----GGVAITAKDSLGLTITGLGVSTAYVGGQITNGRQGGYITGG 231
Qy 242 SNPGTTN-----TWT-----SNNSNTNSQASQNDIDSLKQLYKPLSQRA- 283
Db 232 DMTGTVGRGAVYTTASAVANTVEEFLGGSRTGGSSAGYGVNIDSYSYISGKIASGNGS 291
Qy 284 -VESDGLVFPQAQTSRTARGVAVPHGDH-----YHEIPYS 318
Db 292 LSETIGTGDTLAHTFAGTDSVGYT-GFHIIKTETNLIAKGKFNSDQYIDXSYSIPEQ 350
Qy 319 QMSLEERIAKIIPLRYSNHNWVPSRPEQSPQPTPEPSPPQAPNLIKITSNSGLVQ 378
Db 351 DNEEIKKRL-----QSAHQQLQEOSP----- 371
Qy 379 LVKVGEGVFPEKGISRYVFAKDLPESTVKULESKLSKQSVSHTLTAKKENVAPRDOE 438
Db 372 -----AIYQSKSEDLKALDDEVIRNTLKEHQIQEENDIGQOQE 412
Qy 439 FYDKAYNL-LTEAHKALFYNKGRNSDFQALDKLRLNDESTNKEKLVDDLAFIAPIT 497
Db 413 --DKQLIDLQVREGLYKNQ--QDLK-----QEKRANQOELI----- 446
Qy 498 PRLKPKNSQIETDEVRIAQLADKYTSDGYTDEHDIIISDEGDVAVTPHMGSHWIG 557
Db 447 -----NYELNLQEDQOYELLDO-----LYDEQO----- 471
Qy 558 KDSLSDKEKVAQAQATYKKGILPSPDADVKANPTGDSAAAIYNRVKGEKRIPLVRLPY 617
Db 472 ---QOPQVSNKKQQLQEQQINSPE-----IQYQ 497
Qy 618 VEHTEVKNKGNLIIPIKHYYHNIKPAWPDHTYKAPNGYTLDELPAITIKYVEHNDERP 677
Db 498 LNHLMQ-----PFQDYHNDQTEBELKD-----DDYNFNDQQINNGQFENNVEFPD 543
Qy 678 SNDGKGNASEHYLGKKDSEDPNKAFFKADEEVEESTPAEPVPOVETKEVQQLKEAVL 737
Db 544 LINDADNFEQVANNNNNNNNNNNNNNNFVYDKS-----KSAQOQVEALNBERLY 591
Qy 738 LAKYVDSLIKAVATETLGLRNLTLQIMDNNSIMAEEKLALLKGSPPSSVSKKI 795
Db 592 LOEVED-----APRLIIEIHSNL-----NKAVQEAIEIRQONGNSPAVNSHKI 638

RESULT 21
TANA XENLA
ID TANA XENLA STANDARD; PRT; 1744 AA.
AC Q0150;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tanabin.

```

OS Xenopus laevis (African clawed frog).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
 CC Xenopodinae; Xenopus.
 OK NCBI_TaxID=8355;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC TISSUE=Tadpole head;
 RX MEDLINE=92398961; PubMed=1524825;
 RA Hemmati-Brivianlou A., Mann R.W., Harland R.M.;
 RT "A protein expressed in the growth cones of embryonic vertebrate
 RL Neuron 9:417-428(1992)."
 CC - TISSUE SPECIFICITY: GROWTH CONES OF EMBRYONIC VERTEBRATE NEURONS.
 CC - DEVELOPMENTAL STAGE: IS EXPRESSED IN THE NEURULA AND PERSIST
 CC DURING EMBRYOGENESIS IN THE BRAIN, CRANIAL NERVES, AND SPINAL
 CC CORD.
 CC - SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, M99387; AAA94966.1; -
 DR 21R; JH0720; JH0720.
 DR InterPro; IPR001664; 1F.
 DR Pfam; PF00038; filament; 1.
 DR PROSITE; PS00226; 1F; 1.
 KW Intermediate filament; Coiled coil; Neurone.
 FT DOMAIN 1 12 HEAD.
 FT DOMAIN 13 314 ROD.
 FT DOMAIN 315 1744 TAIL.
 FT DOMAIN 8 48 COIL 1A.
 FT DOMAIN 49 60 LINKER 1.
 FT DOMAIN 61 156 COIL 1B.
 FT DOMAIN 157 179 LINKER 12.
 FT DOMAIN 180 193 COIL 2A.
 FT DOMAIN 194 199 LINKER 2.
 FT DOMAIN 200 314 COIL 2B.
 FT DOMAIN 1744 AA; 199561 MW; 6502EAC9F86C4E93 CRC64;
 SQ SEQUENCE
 Query Match 3.4%; Score 142; DB 1; Length 1744;
 Best Local Similarity 17.1%; Pred. No. 6.7;
 Matches 144; Conservative 145; Mismatches 308; Indels 244; Gaps 37;
 Oy 8 OARTYKENVRSYIIOGKATOKTENLTPEVSKR-----EGINAE-QIVIKI---TDQ 56
 Db 659 KSRKVFLENEYIPVSKDLTEFTSHLENDSESSQFDSLFPNKSTEDLIITNLNSNQE 718
 Oy 57 GVTYSHGDHYHYNYKVPYDAII-----SEELMKDPN-----YKLKDEDIYN 99
 Db 719 NIFQSNQEHLE---NLRFDSVVPPTVKFMYQENNLLEENVYGDELVGATDENIIN 774
 Oy 100 EVKGGYIVKVDGKYVYVYLKDAHADNVTKEI---NRQKEHSHOHRGGTTPRDGAVA 155
 Db 775 Q-----SSDOLLSDSHSHHEETKTESIAVEHNMESEHAE---VDKSE 816
 Oy 156 LARSGRYTTDDGYIFNADSIDIEDGDAYIVPHGHYHIIPNGELASGLAAEAFLSGR 215
 Db 817 IPVEISENVSVETIHEISDVEDITKQAF----- 845
 Oy 216 GNLNSRTYRRQNSDNTSRITNVVPSVNSPGTINTNTSNNSNTNSQASQNDIDSLLKQY 275
 Db 846 ---EDERAGEQINQNNQEST-----VDLDGSVYSGEENSQLEDEVSISEQIEKDFE--- 894
 Oy 276 KLPLSQRHVESDGL--VPDPAQITSRITARGAVAPHGDRHYHPIPYQMSLELRIRIIP 333
 Db 895 --INQECIKKSDQIIEAFDTEEV-----DH-QVVDPMQEOQSFEFEVQQLNNI 938

Oy 334 R-----YRSNMHWPDSPRQPSPOPTPEPSPGOPAPNLKIDNSSLVSYQVRYKVGEGYVF 389
 Db 939 KQEVYILQNY-----DEDSFQNNDEPQ-----ELBS-CDLQEQKIKLEBENGLS 981
 Oy 390 EEKGISRY-----VPAK-----DLPEETKYL-----ESLKSQESVSTL 425
 Db 982 ENEGQNFNGNDIEBFSGQGYDTDEICQETIGNOVSAQLCESDINDOKLMEDEEQNN 1041
 Oy 426 TAKKENVAPR-----DOEFYDKAVNLLTEAHKALFXKGRNSDFQALD- 468
 Db 1042 PETEDNIGLEBESQDENTRSNEGKTFSGQECDDVYFKPEDMDSKESYSGQEDLDKQYTF 1101
 Oy 469 -----KLIER-----LNDESTNKEKLVDDL-AFLADITPERLGKPN 506
 Db 1102 SLNEQANNLDLEKEVILHHDQDQSRVDEITIDKEKSERIIDELATVDVNESLAANKE 1161
 Oy 507 QIETDEVRIFAQL-----ADKYTTSDGIFPEBHDILSDEGDAYVTPHNGHSWIGKD 559
 Db 1162 QVDLFTDEYAVDDVNVGMDDDSGOYOTKEDLFVGGNNI-----EKIEIOQT 1208
 Oy 560 SLSDKEKVAQAAYTKKGLIPSPDADYKANPTGSAAYINRYKGEKRI-PLVRLPYMV 618
 Db 1209 SLNQE-ICERVNDVDEDISGEAKNESVEMADVDLVPEA--KVTGDSQISPLQDEKLN 1265
 Oy 619 EHTVEK--NQNLIIPKHQHYNIKFA---WPDHTYKAPNGYTLDELFAITIKYVEHP 672
 Db 1266 ETMEDTKNDQQLCLEKENETREYIEVTDSPQFATDLSHDARELITVDNSANLQ-FCENP 1324
 Oy 673 DERPHSNDGKGNASEHYLGKDHSEDPNKNFKADEPEPTE---PAPEVQVETIEKVEA 729
 Db 1325 TKT-----LIHNI---EYETVADSDLESTEEQVQETRIIPFKPEDSKEMENSES 1372
 Oy 730 Q 730
 Db 1373 E 1373
 RESULT 22
 MOT3_YEAST
 ID MOT3_YEAST STANDARD; PRT; 490 AA.
 AC P54785;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Zinc finger protein MOT3/HMS1
 GN MOT3 OR HMS1 OR YMR070W OR YN9916.09
 OS Saccharomyces cerevisiae (Baker's yeast).
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 CC NCBI_TaxID=4932;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c;
 RA Madison J., Winston F.;
 RL Submitted (APR-1995) to the EMBL/Genbank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288c / AB972;
 RX PubMed=9169872;
 RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
 RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
 RA Jacobs K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
 RA Rice P., Skellon J., Walsh S., Whitehead S., Barrall B.G.;
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome
 RL XIII."
 RT Nature 387:90-93(1997).
 CC - SUBCELLULAR LOCATION: Nuclear (Potential).
 CC - SIMILARITY: Contains 2 C2H2-type zinc fingers.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way


```

FT CONFLICT 343 343 N -> S (IN REF. 1 AND 3).
FT CONFLICT 421 426 OOKRFL -> TKLSGL (IN REF. 1).
FT CONFLICT 515 515 D -> S (IN REF. 1).
FT CONFLICT 529 535 SKGPRPG -> ARGHDR (IN REF. 1 AND 3).
FT CONFLICT 541 541 D -> V (IN REF. 1 AND 3).
FT CONFLICT 550 551 TD -> LM (IN REF. 1).
FT CONFLICT 573 573 R -> A (IN REF. 1).
FT CONFLICT 582 582 H -> D (IN REF. 1).
FT CONFLICT 588 582 EYTVGWLSSK -> NTLMKAGYPKT (IN REF. 1).
FT CONFLICT 599 599 MISSING (IN REF. 3).
FT CONFLICT 627 632 EKSSSA -> GKULVC (IN REF. 1 AND 3).
FT CONFLICT 635 645 R -> S (IN REF. 1 AND 3).
FT CONFLICT 736 742 ENSTTIT -> RKFMHD (IN REF. 3).
FT CONFLICT 756 756 E -> R (IN REF. 1 AND 3).
FT CONFLICT 773 784 NTKLFKAGVLA -> ILTVFQKLEYMS (IN REF. 1).
FT CONFLICT 793 794 KL -> NV (IN REF. 1).
FT CONFLICT 806 806 N -> T (IN REF. 1).
FT CONFLICT 896 896 N -> NSQITKINTITPOSTYIGERPKRVICGN (IN REF. 1).
FT CONFLICT 906 906 N -> I (IN REF. 1).
FT CONFLICT 911 911 N -> K (IN REF. 1).
FT CONFLICT 915 930 NESLNRYKTSSETIQ -> RIAKILKPAINIT (IN REF. 1).
FT CONFLICT 934 939 DDLVSE -> MTLFL (IN REF. 1).
FT CONFLICT 951 953 AON -> RKT (IN REF. 1).
FT CONFLICT 955 958 BEAH -> KKLD (IN REF. 1).
FT CONFLICT 1002 1002 S -> C (IN REF. 1).
FT CONFLICT 1049 1049 L -> D (IN REF. 1).
FT CONFLICT 1056 1056 C -> S (IN REF. 1).
FT CONFLICT 1060 1060 M -> I (IN REF. 1).
FT CONFLICT 1085 1085 A -> E (IN REF. 1).
FT CONFLICT 1123 1123 V -> C (IN REF. 1).
FT CONFLICT 1133 1133 L -> S (IN REF. 1).
FT CONFLICT 1144 1146 KSN -> NLI (IN REF. 1).
FT CONFLICT 1159 1168 RETKEOBRK -> TRKKEOBRK (IN REF. 1).
FT CONFLICT 1179 1181 SKI -> ELKV (IN REF. 1).
FT CONFLICT 1184 1185 LE -> WK (IN REF. 1).
FT CONFLICT 1188 1204 LSOEISLNQYLNKRISG -> CHRKYLSILKOKNIR (IN REF. 1).
FT CONFLICT 1224 1224 P -> S (IN REF. 1).
FT CONFLICT 1228 1228 E -> Q (IN REF. 1).
FT CONFLICT 1253 1253 E -> Q (IN REF. 1).
FT CONFLICT 1311 1323 PKRESIDINKMLE -> LTKSILITNGNAS (IN REF. 1).
FT CONFLICT 1400 1400 D -> H (IN REF. 1).
FT CONFLICT 1454 1554 SEQULPRLOKDESTEROKELSSSTIKOOKOQFENCMDLQO
NEILRLREHIALKQAEEDVKNNASIIETKTKONKQKELIWI
BREMERNDSDMOQLETL -> P (IN REF. 1).
FT CONFLICT 1568 1568 D -> V (IN REF. 1).
FT CONFLICT 1630 1646 DELKOLDHYTKVEMKLN -> SEARSILYKSGAVD (IN REF. 1).
FT CONFLICT 1688 1704 MISSING (IN REF. 1).
FT CONFLICT 1725 1737 TLQLOMEQNSRNG -> NITANGTKKEM (IN REF. 1).
FT CONFLICT 1754 1757 FDDE -> LMM (IN REF. 1).
FT CONFLICT 1777 1777 D -> E (IN REF. 1).
FT CONFLICT 1788 1788 R -> T (IN REF. 1).
FT CONFLICT 1825 1825 S -> D (IN REF. 1).
FT CONFLICT 1882 1882 S -> W (IN REF. 1).
FT CONFLICT 1902 1904 FMK -> NSGKLLDLDL (IN REF. 1).
SQ SEQUENCE 1928 AA; 221634 MW; 6F54C7611FA3DC9F CRC64;

Query Match 3 4*: Score 141.5; DB 1; Length 1928;
Best Local Similarity 19.6%; Pred. No. 8.1;
Matches 174; Conservative 132; Mismatches 318; Indels 265; Gaps 41;

```

```

QY 104 GYVTKVDGKYVYVLKDAAHADNVRTKEINRQOEHOSHREGGTFRNDGAVALAROGRY 163
DB 1094 -----KESBLKMKKEIKDHNKKEIATF-----SKORDAYS-----EHKXI 1129
QY 164 TTDDGYVFNASDIEDTGDAVYVPHGDHYHYIPXNELSASELAAAEAFLSGRNUL----- 218
DB 1130 TAB-----LKET-RIOLTEYKSNYQKI-KEEYSNFORETEOROKRNSLVESLN 1177
QY 219 -SNSRTYRQNSDNTSRNWPVSNSPCTNTNTNNSNTNSQASQNSDIDSLKOLYKL 277
DB 1178 DSKIKELERARISQISLNOYLANKRISGNSVETNISSTRSISYSDDPLDKEDITIKKYDL 1237
QY 278 PLS-----QRHVESD-----GLVEPDAQITRTARGAVVAPGDHYHFIYPSQSELE 324
DB 1238 QLATETITRNLENLEIEKKNLISRLRFTERTLASSPEQDQIK-----AQMKLX 1287
QY 325 ERIARIIPLRYSNHWVPDSRPEQSPQPRPEPSGPQAPNLKIDSNSL--VSQVLRK 382
DB 1288 KLI-----QMDPSIPLDSILNEPLDN---CPDKESDIKMLLEVYDLKRQ 1330
QY 383 VG---EGVFEKGISR-----YFADLSEYTKNLESTKSQES 420
DB 1331 LDIBTRAHYDAENALISALHSXFRKIQGESSLSSSDIYKLFKEASEERVKSLIEDLKTMP 1390
QY 421 VSHF-----LTAKKENVAPRDOE--FYD-----KAYNLTLEAHKAL-----FXNKGNSDF 464
DB 1391 RDRTNLPVPGDIIKRRDSISKYEEIRYKLENYLQELINSNGKLSQTLIDLQOSKKE 1450
QY 465 QALDKLERLND--ESTNKEKLVDDLAFAPITHPERLGRKPSQIEYTEDEVRIQALAD 522
DB 1451 ALLSEQDRLQKDXLESTERQ--ELLS-----STIKOOKOQFENCMDLQONEIR- 1498
QY 523 KYTTSQGIYFPEHDIIDEGGAVYTPPHMGSHWIGKDSLQSKEXVAAQATYKKGILP-- 580
DB 1499 -----LREHIALKQAEEDV-----KNNASIIETKTKONKQKELIWERE 1538
QY 581 -PSPDADYKANPTGDSAAAIYNRVKGKRIPLVRLPYMVEHTEVYKONLIIIPKHNYHN 639
DB 1539 MERNDSDMOQLETL-----LLEIKRVQDYKXKILSDDLAHLKELSAVEDS---QYTDENR 1591
QY 640 IKFAMPDHYTKABNGYLTLEDLFAITIKYVVEHPDERPHSNDGWNASEHVLGKKDHSEDP 699
DB 1592 LK-----BELNCSLKAETNLKKEFATLYKLE-----TSTNDSKAKISD-LLKQDHYTKV 1641
QY 700 NKNPKADEEPEVEFPAPREYVQVETEKYEALKAEBVLAV-----T 742
DB 1642 VEMLNNEKDAL--SLAEKELYQ---KYBALNTECESLKGKIVSLTYKIQELESDLNOKT 1695
QY 743 DSSLKANA-----TETLAGLRNNLTLOIMDNNSIMAEAEKLL 779
DB 1696 DALQISNBAISSLSTQKKEITEKIKYLEETLQLO-MEONSGNGLVTKL 1743

RESULT 24
PTPZ_HUMAN
ID_PTPZ_HUMAN STANDARD; PRT: 2314 AA.
AC P23477.1
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein-tyrosine phosphatase zeta precursor (BC 3.1.3.48) (R-PTP-
DE zeta).
GN PTPRZ1 OR PTPRZ OR PTPZ.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM LONG).
RC TISSUE=Brain;
RX MEDLINE=92366472; PubMed=1323835;
RA Krueger N.X.; Saito H.;

```

RT "A human transmembrane protein-tyrosine-phosphatase, PTP zeta, is
RT expressed in brain and has an N-terminal receptor domain homologous
RT to carbonic anhydrases."
RT Proc. Natl. Acad. Sci. U.S.A. 89:7417-7421 (1992).
RN (2)
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RC TISSUE=Brain stem;
RX MEDLINE=9352948; PubMed=8387522;
RA Levy J.B., Canoll P.D., Silvenoinen O., Barnea G., Morse B.,
RA Honegger A.M., Huang J.-T., Cannizzaro L.A., Park S.-H., Druck T.,
RA Huebner K., Sap J., Ehrlich M., Musacchio J.M., Schlesinger J.,
RT "The cloning of a receptor-type protein tyrosine phosphatase
RT expressed in the central nervous system."
RT J. Biol. Chem. 268:10573-10581 (1993).
RN (3)
RP SEQUENCE OF 1479-2091 FROM N.A. (ISOFORM LONG).
RC TISSUE=Liver;
RX MEDLINE=9106018; PubMed=2170109;
RA Krueger N.X., Streuli M., Saito H.,
RT "Structural diversity and evolution of human receptor-like protein
RT tyrosine phosphatases."
RL EMO J. 9:3241-3252 (1990).
RN (4)
RP SEQUENCE OF 1749-1990 AND 2047-2280 FROM N.A. (ISOFORM LONG).
RC TISSUE=Brain stem;
RX MEDLINE=90384936; PubMed=2169617;
RA Kaplan R., Morse B., Huebner K., Croce C., Hawk R., Ravera M.,
RA Ricca G., Jaye M., Schlesinger J.,
RT "Cloning of three human tyrosine phosphatases reveals a multigene
RT family of receptor-linked protein-tyrosine-phosphatases expressed in
RT brain."
RL Proc. Natl. Acad. Sci. U.S.A. 87:7000-7004 (1990).
CC -1- FUNCTION: MAY BE INVOLVED IN THE REGULATION OF SPECIFIC
CC DEVELOPMENTAL PROCESSES IN THE CNS.
CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O = protein
CC tyrosine + phosphate.
CC -1- SUBUNIT: THE CARBONIC-ANHYDRASE LIKE DOMAIN BINDS TO CONTACTIN (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P23471-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P23471-2; Sequence=VSP_005151;
CC -1- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN THE CENTRAL NERVOUS
CC SYSTEM, WHERE IT IS LOCALIZED IN THE PURKINJE CELL LAYER OF THE
CC CEREBELLUM, THE DENTATE GYRUS, AND THE SUBENDYMEAL LAYER OF THE
CC ANTERIOR HORN OF THE LATERAL VENTRICLE. DEVELOPMENTALLY REGULATED
CC IN THE BRAIN.
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, BELONGS TO THE EUKARYOTIC-
CC TYPE CARBONIC ANHYDRASE FAMILY.
CC -1- SIMILARITY: Contains 2 protein-tyrosine phosphatase domains.
CC -1- SIMILARITY: Contains 1 fibronectin type III domain.
CC -1- CAUTION: CALLED RPTPASE BETA IN REF.2 AND REF.4.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
DR EMBL: M93426; AAA60225.1; -
DR EMBL: X54135; CAA38070.1; -
DR PIR: A46151; A46151.
DR HSSP: P18052; LYFO.
DR Genew: HGNC:9665; PTPRZ1.
DR MIM: 176891; -
DR GO: GO:0005887; C: integral to plasma membrane; TAS.
DR GO: GO:0005001; P: transmembrane receptor protein tyrosine pho. .; TAS.
DR GO: GO:0007417; P: central nervous system development; TAS.

DR GO: GO:0006470; P: protein amino acid dephosphorylation; TAS.
DR InterPro: IPR001148; Euk Coahd.
DR InterPro: IPR003961; FN III.
DR InterPro: IPR000387; TYR phosphatase.
DR InterPro: IPR000242; TyP.
DR Pfam: PF00194; carb_anhydrase; 1.
DR Pfam: PF00041; fn3; 1.
DR Pfam: PF00102; Y phosphatase; 2.
DR PRINTS: PR00700; PRTYHPHPTASE.
DR Prodom: PD000865; Euk Coahd; 1.
DR SMART: SM00060; FN3; 1.
DR SMART: SM00194; PTPC; 2.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
DR PROSITE: PS00056; TYR_PHOSPHATASE_2; 2.
DR PROSITE: PS00055; TYR_PHOSPHATASE_PTP; 2.
KW Signal: Glycoprotein; Transmembrane; Hydrolase; Repeat;
KW Alternative splicing.
FT CHAIN 1 24
FT SIGNAL 1 24
FT CHAIN 25 2314
FT DOMAIN 25 1635
FT TRANSMEM 1636 1661
FT DOMAIN 1662 2314
FT DOMAIN 34 302
FT DOMAIN 312 406
FT DOMAIN 1744 1997
FT DOMAIN 1998 2314
FT ACT_SITE 1932 1932
FT SITE 2222 2222
FT CARBOHYD 105 105
FT CARBOHYD 134 134
FT CARBOHYD 223 223
FT CARBOHYD 232 232
FT CARBOHYD 334 324
FT CARBOHYD 381 381
FT CARBOHYD 497 497
FT CARBOHYD 501 501
FT CARBOHYD 552 552
FT CARBOHYD 587 587
FT CARBOHYD 602 602
FT CARBOHYD 629 629
FT CARBOHYD 637 637
FT CARBOHYD 677 677
FT CARBOHYD 997 997
FT CARBOHYD 1017 1017
FT CARBOHYD 1050 1050
FT CARBOHYD 1082 1082
FT CARBOHYD 1122 1122
FT CARBOHYD 1456 1456
FT CARBOHYD 1548 1548
FT CARBOHYD 1550 1550
FT CARBOHYD 1561 1561
FT CARBOHYD 1617 1617
FT VARSPLIC 755 1614
FT CONFLICT 1722 1728
SQ SEQUENCE 2314 AA; 254528 MW; 77DBDEFA05FB42 CRC64;
Query Match 3.4%; Score 141.5; DB 1; Length 2314;
Best Local Similarity 19.2%; Pred. No. 10;
Matches 169; Conservative 113; Mismatches 267; Indels 331; Gaps 43;
QY 51 IKITDQGVYTSHGDIHYHYNGKVPYDAIISEELMKDKNYKDKBDIVNEVGVIKYD 110
DB 954 VGVTYQGLFSGPSRII-----PIPKSLITPASILOPHTALSG-----ARSQGRYT 164
DB 111 GKYVYVYLDAADNVRTKEINRQKHSHOHREGGTPRNDCAVAL-----ARSQGRYT 164
DB 994 GEM-----SGASSDS-----EFLPDTGLTALNISPVSVAEFYVT 1030

```

OY 165 T---DDGYFNASDIIEDTGDAYIVPHGDHYH-----IPK-----NELSASELAAB 209
DB 1031 TSVFDDDKALKSKSEIITGNTELOIPSENVVSESEVMNMNDNNKLNASIQEISV 1090
OY 210 AFLSGRG-----NLSNRTYR-RQNSDNTS-----RTNVVP 239
DB 1091 SSSSTKMGFPGSLAHTTKVPFHEISQVPENNFVSQVPHTVASQASGDTSLKPVLSANSEP 1150
OY 240 SVSNGGT-----NNTSNNTNNTNQS-ONSDDISLLKQLYKPLSGRHVES 286
DB 1151 ASSDPASSEMUSPSTOLLFFETSAFSTEVLLQPSFQASDVTLTKTV--LPA---VPS 1204
OY 287 DGLVDPDPAQ-----ITSRTARGVAVPHGDHYHIFPKSQMSELEERARIIPRY 335
DB 1205 DPLIVETKRVKDKISSTMLHLIVSNASSENMTHTSVFVDVSPSHMSASLOGLTISY 1264
OY 336 RSNHWVPDSRPQSPQPTPEPSPGPQAPAPNLKIDSNSLSVSLVRYKVGEGVFEKGIS 395
DB 1265 ASEKYEPLVLKSGSSHQVVP-----SLVSNDELFOQAMLEINQAH--PPKG-- 1308
OY 396 RYVFAKDLPS--ETVKANLESKSKQESVSHLTAKKENVAPR-----DQEFYDAX 444
DB 1309 RHVFATPVLSIDEPNTLTKLHSDER--LTSKTSVTKGVAGIPTVASDTRFVSTDH 1365
OY 445 NL-LTEAH-----KALFXNKG----- 459
DB 1366 SVPIKNGVAVTAVPVRHDSVTSKTLFSPSATSELSAKSDAGLVGGEGDGTDDG 1425
OY 460 -----RNSDFOALDKLERLNDSTNKEKLVDDLFLAPITPRLGKPNISLEY----- 510
DB 1426 DDDDRDRSDGSLHCKMS-CSSYRESQEKVMD-----SDTHENSLMDQNNPISYLSSE 1478
OY 511 -TEDEVRAQLADKTYT-----SDGYITDEHDI-----ISDEG 542
DB 1479 NSEEDNRVTVSSDSQTMDRSPGKSPSANGLSQKHNDDK--EENDIQTSALLPLSPBS 1536
OY 543 DAYVTPPHMGSHWIK-----DSLSDKEKVA--AQAYTKER--GIY-----PSP 583
DB 1537 KAMAVLTDSESGSQGSDSLNENETSDPSFADTNEGDAGIILAGDSEITTPGPSP 1596
OY 584 DADV-----KANPTGDSAAIYVRKGEKR--IPVLVLPYWEHTEVYKNG 627
DB 1597 TSSVTSENSEVPHVSEAEASNSHRSRIGLAGLSEKKAIVPLIVSALFTICVLVWG 1656
OY 628 NL-----IIPKDHNTKPFAMPDHT--YQAPNGY 656
DB 1657 ILIYWRKCFQTAHFYLEDSTSPRVISTPPTPIFPISDVGALPIKHFPGVADLHASG 1716
OY 657 TLEDLFATIK-YVVE-----HDERPHSN 679
DB 1717 TEE--FETIKERYQEVQCTVDLGITADSSNHPDK-HKN 1753

```

```

RT "Nuclear migration in Saccharomyces cerevisiae is controlled by the
RL highly repetitive 313 kDa NUM1 protein."
DB Mol. Gen. Genet. 230:277-287(1991).
OY -1- FUNCTION: CONTROLS NUCLEAR MIGRATION. NUM1 SPECIFICALLY CONTROLS
DB THE INTERACTION OF THE BUD NECK CYTOSKELETON WITH THE PRE-
OY DIVISIONAL G2 NUCLEUS. PERHAPS BY RECOGNIZING G2-SPECIFIC
DB CYTOPLASMIC MICROTUBULI OR OTHER COMPONENTS OF THE NUCLEAR
OY ENVELOPE.
DB -1- MISCELLANEOUS: ADDITIONAL REGIONS OF LOWER HOMOLOGY TO THE REPEAT
OY CONSENSUS (ALWAYS STARTING WITH PROLINE) ARE FOUND IN BOTH
DB FLANKING DOMAINS OF THE TANDEM REPEATS.
OY -1- SIMILARITY: Contains 1 PH domain.
DB -1- CAUTION: IT IS UNCERTAIN WHETHER MEN-1 IS THE INITIATOR.
OY -----
DB This SWISS-PROT entry is copyright. It is produced through a collaboration
OY between the Swiss Institute of Bioinformatics and the EMBL outstation
DB the European Bioinformatics Institute. There are no restrictions on its
OY use by non-profit institutions as long as its content is in no way
DB modified and this statement is not removed. Usage by and for commercial
OY entities requires a license agreement (See http://www.isb.ch/announce/
DB or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X61236; CAA4354.1; -.
CC SCD: S0002557; NUM1.
CC InterPro: IPR001849; PH.
CC Pfam: PF00169; PH; 1.
CC SMART: SMO0233; PH; 1.
CC PROSITE: PSS0003; PH_DOMAIN; 1.
CC Repeat.
CC DOMAIN
CC REPEAT 593 1384 12.5 X TANDEM REPEATS.
CC REPEAT 593 656 1.
CC REPEAT 657 727 2.
CC REPEAT 728 798 3.
CC REPEAT 799 862 4.
CC REPEAT 863 926 5.
CC REPEAT 927 990 6.
CC REPEAT 991 1054 7.
CC REPEAT 1055 1118 8.
CC REPEAT 1119 1182 9.
CC REPEAT 1183 1246 10.
CC REPEAT 1247 1310 11.
CC REPEAT 1311 1374 12.
CC REPEAT 1375 1384 13 (INCOMPLETE).
CC DOMAIN 2573 2683 PH.
CC SEQUENCE 2748 AA; 313202 MW; B2FBD67C9F6211AE CRC64;

```

Query Match 3.4%; Score 141.5; DB 1; Length 2748;
Best Local Similarity 18.8%; Pred. No. 13;
Matches 151; Conservative 124; Mismatches 315; Indels 213; Gaps 34;

```

OY 70 NGKVPYDAIIESEELMKDPNKLKDEDIYNVKGQYVKKVGDQYVYLKDAHADNVRTK 129
DB 769 SAKLGHTVSSSEYSELOKTSLEKEVEQPSLAYVEHAQATDHHLLSDAYEELVCK 828
OY 130 EFINROKQSHQHEGTPRNDGAVALARSGQRYTTDGYIFNASDIIEDTGDAYIVPHG 189
DB 829 ENPMPEFLKESAKLGHTVVSNAYSELEK-----LEQPSLAYVEHA 872
OY 190 ---DHYHIIIPNLSASELAAEAPFLSGRGLNSRTYR-----RQNSDNTSRINWVSV 241
DB 873 KATDHH-----LLSDSAYEDLVCKENSD-----VFL 900
OY 242 SNGPTNTNNTSNNTNNTNQSQNSNDISLLKQLYKPLSGRHVESDGLVFPDPAQ----- 295
DB 901 KEKSAKLGHTVVSNAYSELEKLEQPSLAYVEHAQATDHHLLSDAYEDLVCKENPD 960
OY 296 ---ITSRTAR-GVAVPHGDHYHIFPKSQMSELEERAR-----IIPRYSNHWVPDSR 345
DB 961 MEFLKEKSAKLGHTVVSNAY-----SLEKKLQPSLEYVEHAQATDHHLLSDSA 1012
OY 346 -----PEQSPQPTPEPSP--GPQAPAPNLKIDSNSLSVSLVRYKVGEGVFEKGISR 396
DB 1013 YEDLVCKENPDMEFLKESAKLGHTVVSN---EAYSLEKLEQPSLEYVEHAQATNH 1069

```

```

OY 397 YVFANDLPSETVK-----NLESKLSKQESVSHLTAKENNVAPRDOEFYDKAYNLLTEAH 451
D 1070 HLLSDSAVELVKCKENPDVEFLKESAKLGHVVV--NEAYSLEKKLPSLEVEVHA 1128
OY 452 KALFANKGNSDFQALDKLLERLNDSSTNKEKLVDDLAFIAPITPERLGKPNQIERYT 511
D 1129 KATNHLILSDSAVEELVKCKEN--PDVEFLKESAKLGHVVV--NEAYSLEKK 1179
OY 512 EDEVRALADKVTTSDDGYIFDEHDIIISDEGDVAVTPHMGHSHWIKDLSIDSEKVAQA 571
D 1180 LEQPSLAYVEHAKATD-----HHLISD--SAV-----EDLVCKENPDVE-- 1218
OY 572 YTKKCGILPPSPDADVKANPTGDSAAAIYNRVGEKRIPLVLPVNEHTVEVKGNLLI 631
D 1219 FLKEK-----SAKLGHVVV--NEAYSLEKKLPSLEVEVHA--PSL 1259
OY 632 PKHDYHNIFKAFMD-----HTYKAPNGYT-----LEDIFAT 665
D 1260 ---DH--HLLSDSAVELVKCKENPDVEFLKESAKLGHVVV--NEAYSLEKKLEQ-- 1313
OY 666 KYVVEHPDERPHNSDQMGNASEHVLGKDH-----SEDPNKNF-----KADEEVEET 713
D 1314 EYLVEHAKATNH-----HLLSDSAVELVKCKENPDVEFLKESAKLGHVVVSNK 1363
OY 714 PAEPVPEVETEKEVEAQLKEAEVLLAK--VTDSSILKAN--ATEPLAGLNNLTLQIMDN 769
D 1364 EXSELEKLEKPSLEVLVHGAEOIGKTIISIDPNTLANPSEMDMAKLOKLEVOYLSND 1423
OY 770 SIMA-----EAEKLLALLNG 784
D 1424 EYIALKNTEKPEKPDVELLRSLKLG 1446

RESULT 26
TIF42_YEAST STANDARD; PRT; 914 AA.
AC P39936;
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Eukaryotic initiation factor 4F subunit p130 (eIF-4F) (mRNA cap-
binding protein complex subunit p130).
GN TIF4632 OR YGL049C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycace.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93330281; PubMed=8336723;
RA Goyer C., Altman M., Lee H.S., Blanc A., Deshmukh M.,
RA Woolford J.L., Trachsel H., Sonenberg N.;
RA "TIF4632 and TIF4632: two yeast genes encoding the
RT high-molecular-weight subunits of the cap-binding protein complex
RT (eukaryotic initiation factor 4F) contain an RNA recognition
RT motif-like sequence and carry out an essential function.";
RL Mol. Cell. Biol. 13:4860-4874 (1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=6288C;
RX MEDLINE=93737993; PubMed=9234674;
RA Feuermann M., de Montigny J., Potier S., Souciet J.-L.;
RA "The characterization of two new clusters of duplicated genes
RT suggests a 'lego' organization of the yeast Saccharomyces cerevisiae
RT chromosomes".
RL Yeast 13:861-869 (1997).
CC -I- FUNCTION: INTERACTS WITH THE MNA CAP STRUCTURE. MTGPPX. THIS
CC INTERACTION IS REQUIRED FOR EFFICIENT RIBOSOME BINDING TO THE
CC MRNA. TIF4632 IS PROBABLY ESSENTIAL WHEN TIF4631 IS MISSING.
CC -I- SUBUNIT: MAY FORM PART OF THE CAP-BINDING PROTEIN COMPLEX.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L16924; AAA18474.1; -.
DR EMBL; Z72571; CAA96751.1; -.
DR PIR; B48086; B48086
DR SGD; S0003017; TIF4632.
DR InterPro; IPR003890; IF_eIF4G.
DR InterPro; IPR000504; RNA_rec_mot.
DR Pfam; PF02854; MIF4G; 1.
DR SMART; SM00543; MIF4G; 1.
KW Initiation factor; Protein biosynthesis; Translation regulation;
KW RNA-binding; Multigene family.
FT DOMAIN 32 97 ASN-RICH.
FT DOMAIN 459 510 ARG/SER-RICH.
FT DOMAIN 840 863 ARG/SER-RICH.
SQ SEQUENCE 914 AA; 103898 MW; BF5E9805CD47908E CRC64;

Query Match 3.4%; Score 141; DB 1; Length 914;
Best Local Similarity 19.7%; Pred. No. 3.1;
Matches 130; Conservative 107; Mismatches 252; Indels 170; Gaps 34;

OY 187 PHGDHY-----HTPKNELSASELAFAELSGRLNSRTTRRONSDDTSTRTNWPSVS 242
D 23 PHDNOYSGANNQPPNNHYENLYSAREP-----HNKKOYQSKN-----G 61
OY 243 NGPTNTNTSNNNSNQASQNSIDSLKQLYKLPLSGRHVESDGLVDFPAPQITRTAR 302
D 62 KYGTNKKYNNRNNNSQNAQY-----YNNRPNNGYRLNNND-----YNPMLP----- 102
OY 303 GVAVPHGDHYHFIYPSQNSELEERLARIILPRLSYNNHWVDSRPPQSPQTPPEBSPGQ 362
D 103 GMQWPMAN--YYAP--QWYIIPQWVPVAPSPY--THQPLANTNPEPPS---TPKTT---- 148
OY 363 PAPNLKIDNSNSLVQL-----VRKQGEVY-----PEEKISRYVPAKD-----LP-SE 406
D 149 ----KIETTYTGRLLMKKHEEKASKGEEKNDGVEQKSGTPEFEKATPVLPANPE 203
OY 407 TVKNLESKLSKQESVSHLTAK--ENVAAPRDOEFYDKAYVLLTEAHALFXNKCRRNSD 463
D 204 AVKDLITETSNKSTSEALNTGRLLFLEQYRLRKAAMERKKNGLISTEKKQSTSHNDTD 263
OY 464 FOALDKLLERLNDSSTNK-----EKVYDDLAFIAPITPERL--GKPNQIERYTEDEV 516
D 264 TTKPNSVIESEPIKAPKPTGEANEVVIDGKSG--ASVKTPOHVGSVTKSVTFNEPE-- 319
OY 517 IQQLADKYTTSDDGYIFDEHDIIISDEGDVAVTPHMGHSHWIKGD--SLDSKEKYAAQAATKE 575
D 320 -----NESSQDDELVDKDDTTEISDTTG-----GKTVNKSDETETNSVTTTEE 364
OY 576 KGIILPESP-ADV-----KANPTGDSAAAIYNR-----VKGEKRIPLVRLP 615
D 365 NTVKETESTSIEMPTYSQLETKGKAPISDIYEPAYPEVNEVPDIIYKK--DSVKTT 422
OY 616 Y-----MVEHT-----VEVKNGLIIPKHDYHNIFKAFWPDHTYKAPN-----GY 656
D 423 YGPTFLQFKDKLFRPDPAWEAVSISKVIVIPPHIARNKPXSGRFGGFRSPKRGMDH 482
OY 657 TLEDI.FATIKYVVEHPDERPHNSDQMGNASEHVLGKDHSEDPNKNFKADEEVEETPAE 716
D 483 TSSSEVSSKRRKRGMDR--RSNRG-----TSRKORE-----KAAEYAEQAPKE 527
OY 717 ---PEVPEVETEKEVEAQLKEAEVLLAKVTDSS--KANATETLAGLNNLTLQIMDNNS 770
D 528 ETAPLVPSANRWIIPKSRVAKTEKKLAPDGKTELPKEVEVERKMSLNLKLTLEMPDIS 586

RESULT 27
SP7_YEAST

```

ID SPT7 YEAST STANDARD; PRT; 1332 AA.
AC P15177;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Transcriptional activator SPT7.
GN SPT7 OR YBR081C OR YBR0739.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Bacteria; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OK NCBI_TaxId=4932;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=95229044; PubMed=7713415;
RA Gansheroff L.J., Dollard C., Tan P., Winston F.;
RT "The Saccharomyces cerevisiae SPT7 gene encodes a very acidic protein
important for transcription in vivo."
RL Genetics 139:523-536(1995).
RN 21
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=95076715; PubMed=7985423;
RA van der Aart Q.J.M., Barthe C., Daignon F., Aigle M., Crouzet M.,
RA Steensma H.Y.;
RT "Sequence analysis of a 31 kb DNA fragment from the right arm of
Saccharomyces cerevisiae chromosome II."
RL Yeast 10:959-964(1994).
RN 13
RP SEQUENCE OF 1-835 FROM N.A.
RC STRAIN=S288C;
RA Andre B., Czapluch C., Hein C., Jauniaux J.C., Urrestarazu A.,
RA Vissers S.;
RT Submitted (AUG-1994) to the EMBL/GenBank/DBJ databases.
RN 14
RP SEQUENCE OF 463-523 FROM N.A.
RX MEDLINE=92285152; PubMed=1350857;
RA Haynes S.R., Dollard C., Winston F., Beck S., Trowdale J.,
RA David I.B.;
RT "The bromodomain: a conserved sequence found in human, Drosophila and
yeast proteins."
RL Nucleic Acids Res. 20:2603-2603(1992).
CC - FUNCTION: TRANSCRIPTIONAL ACTIVATOR OF TY ELEMENTS AND POSSIBLY
OTHER GENES.
CC - SUBCELLULAR LOCATION: Nuclear.
CC - SIMILARITY: Contains 1 bromodomain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
at the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L22537; AAC37424.1; -
DR EMBL: X75294; CAAS3940.1; -
DR EMBL: Z35950; CA85026.1; -
DR EMBL: M87651; AAA35087.1; -
DR PIR: S41552; S41552.
DR HSSP: Q92831; 1891.
DR TRANSFAC: T04835; -
DR SGD: S0000285; SPT7.
DR GO: GO:0000124; C:SAGA complex; IDA.
DR InterPro: IPR001487; Bromodomain.
DR Pfam: PF00439; Bromodomain; 1.
DR PRINTS: PRO0503; BROMODOMAIN.
DR SMART: SM00297; BROMO; 1.
DR PROSITE: PS00633; BROMODOMAIN_1; 1.
DR PROSITE: PS00633; BROMODOMAIN_2; 1.
DR PROSITE: PS00633; BROMODOMAIN_3; 1.
DR PROSITE: PS00633; BROMODOMAIN_4; 1.
DR PROSITE: PS00633; BROMODOMAIN_5; 1.
DR PROSITE: PS00633; BROMODOMAIN_6; 1.
DR PROSITE: PS00633; BROMODOMAIN_7; 1.
DR PROSITE: PS00633; BROMODOMAIN_8; 1.
DR PROSITE: PS00633; BROMODOMAIN_9; 1.
DR PROSITE: PS00633; BROMODOMAIN_10; 1.
DR PROSITE: PS00633; BROMODOMAIN_11; 1.
DR PROSITE: PS00633; BROMODOMAIN_12; 1.
DR PROSITE: PS00633; BROMODOMAIN_13; 1.
DR PROSITE: PS00633; BROMODOMAIN_14; 1.
DR PROSITE: PS00633; BROMODOMAIN_15; 1.
DR PROSITE: PS00633; BROMODOMAIN_16; 1.
DR PROSITE: PS00633; BROMODOMAIN_17; 1.
DR PROSITE: PS00633; BROMODOMAIN_18; 1.
DR PROSITE: PS00633; BROMODOMAIN_19; 1.
DR PROSITE: PS00633; BROMODOMAIN_20; 1.
DR PROSITE: PS00633; BROMODOMAIN_21; 1.
DR PROSITE: PS00633; BROMODOMAIN_22; 1.
DR PROSITE: PS00633; BROMODOMAIN_23; 1.
DR PROSITE: PS00633; BROMODOMAIN_24; 1.
DR PROSITE: PS00633; BROMODOMAIN_25; 1.
DR PROSITE: PS00633; BROMODOMAIN_26; 1.
DR PROSITE: PS00633; BROMODOMAIN_27; 1.
DR PROSITE: PS00633; BROMODOMAIN_28; 1.
DR PROSITE: PS00633; BROMODOMAIN_29; 1.
DR PROSITE: PS00633; BROMODOMAIN_30; 1.
DR PROSITE: PS00633; BROMODOMAIN_31; 1.
DR PROSITE: PS00633; BROMODOMAIN_32; 1.
DR PROSITE: PS00633; BROMODOMAIN_33; 1.
DR PROSITE: PS00633; BROMODOMAIN_34; 1.
DR PROSITE: PS00633; BROMODOMAIN_35; 1.
DR PROSITE: PS00633; BROMODOMAIN_36; 1.
DR PROSITE: PS00633; BROMODOMAIN_37; 1.
DR PROSITE: PS00633; BROMODOMAIN_38; 1.
DR PROSITE: PS00633; BROMODOMAIN_39; 1.
DR PROSITE: PS00633; BROMODOMAIN_40; 1.
DR PROSITE: PS00633; BROMODOMAIN_41; 1.
DR PROSITE: PS00633; BROMODOMAIN_42; 1.
DR PROSITE: PS00633; BROMODOMAIN_43; 1.
DR PROSITE: PS00633; BROMODOMAIN_44; 1.
DR PROSITE: PS00633; BROMODOMAIN_45; 1.
DR PROSITE: PS00633; BROMODOMAIN_46; 1.
DR PROSITE: PS00633; BROMODOMAIN_47; 1.
DR PROSITE: PS00633; BROMODOMAIN_48; 1.
DR PROSITE: PS00633; BROMODOMAIN_49; 1.
DR PROSITE: PS00633; BROMODOMAIN_50; 1.
DR PROSITE: PS00633; BROMODOMAIN_51; 1.
DR PROSITE: PS00633; BROMODOMAIN_52; 1.
DR PROSITE: PS00633; BROMODOMAIN_53; 1.
DR PROSITE: PS00633; BROMODOMAIN_54; 1.
DR PROSITE: PS00633; BROMODOMAIN_55; 1.
DR PROSITE: PS00633; BROMODOMAIN_56; 1.
DR PROSITE: PS00633; BROMODOMAIN_57; 1.
DR PROSITE: PS00633; BROMODOMAIN_58; 1.
DR PROSITE: PS00633; BROMODOMAIN_59; 1.
DR PROSITE: PS00633; BROMODOMAIN_60; 1.
DR PROSITE: PS00633; BROMODOMAIN_61; 1.
DR PROSITE: PS00633; BROMODOMAIN_62; 1.
DR PROSITE: PS00633; BROMODOMAIN_63; 1.
DR PROSITE: PS00633; BROMODOMAIN_64; 1.
DR PROSITE: PS00633; BROMODOMAIN_65; 1.
DR PROSITE: PS00633; BROMODOMAIN_66; 1.
DR PROSITE: PS00633; BROMODOMAIN_67; 1.
DR PROSITE: PS00633; BROMODOMAIN_68; 1.
DR PROSITE: PS00633; BROMODOMAIN_69; 1.
DR PROSITE: PS00633; BROMODOMAIN_70; 1.
DR PROSITE: PS00633; BROMODOMAIN_71; 1.
DR PROSITE: PS00633; BROMODOMAIN_72; 1.
DR PROSITE: PS00633; BROMODOMAIN_73; 1.
DR PROSITE: PS00633; BROMODOMAIN_74; 1.
DR PROSITE: PS00633; BROMODOMAIN_75; 1.
DR PROSITE: PS00633; BROMODOMAIN_76; 1.
DR PROSITE: PS00633; BROMODOMAIN_77; 1.
DR PROSITE: PS00633; BROMODOMAIN_78; 1.
DR PROSITE: PS00633; BROMODOMAIN_79; 1.
DR PROSITE: PS00633; BROMODOMAIN_80; 1.
DR PROSITE: PS00633; BROMODOMAIN_81; 1.
DR PROSITE: PS00633; BROMODOMAIN_82; 1.
DR PROSITE: PS00633; BROMODOMAIN_83; 1.
DR PROSITE: PS00633; BROMODOMAIN_84; 1.
DR PROSITE: PS00633; BROMODOMAIN_85; 1.
DR PROSITE: PS00633; BROMODOMAIN_86; 1.
DR PROSITE: PS00633; BROMODOMAIN_87; 1.
DR PROSITE: PS00633; BROMODOMAIN_88; 1.
DR PROSITE: PS00633; BROMODOMAIN_89; 1.
DR PROSITE: PS00633; BROMODOMAIN_90; 1.
DR PROSITE: PS00633; BROMODOMAIN_91; 1.
DR PROSITE: PS00633; BROMODOMAIN_92; 1.
DR PROSITE: PS00633; BROMODOMAIN_93; 1.
DR PROSITE: PS00633; BROMODOMAIN_94; 1.
DR PROSITE: PS00633; BROMODOMAIN_95; 1.
DR PROSITE: PS00633; BROMODOMAIN_96; 1.
DR PROSITE: PS00633; BROMODOMAIN_97; 1.
DR PROSITE: PS00633; BROMODOMAIN_98; 1.
DR PROSITE: PS00633; BROMODOMAIN_99; 1.
DR PROSITE: PS00633; BROMODOMAIN_100; 1.
DR PROSITE: PS00633; BROMODOMAIN_101; 1.
DR PROSITE: PS00633; BROMODOMAIN_102; 1.
DR PROSITE: PS00633; BROMODOMAIN_103; 1.
DR PROSITE: PS00633; BROMODOMAIN_104; 1.
DR PROSITE: PS00633; BROMODOMAIN_105; 1.
DR PROSITE: PS00633; BROMODOMAIN_106; 1.
DR PROSITE: PS00633; BROMODOMAIN_107; 1.
DR PROSITE: PS00633; BROMODOMAIN_108; 1.
DR PROSITE: PS00633; BROMODOMAIN_109; 1.
DR PROSITE: PS00633; BROMODOMAIN_110; 1.
DR PROSITE: PS00633; BROMODOMAIN_111; 1.
DR PROSITE: PS00633; BROMODOMAIN_112; 1.
DR PROSITE: PS00633; BROMODOMAIN_113; 1.
DR PROSITE: PS00633; BROMODOMAIN_114; 1.
DR PROSITE: PS00633; BROMODOMAIN_115; 1.
DR PROSITE: PS00633; BROMODOMAIN_116; 1.
DR PROSITE: PS00633; BROMODOMAIN_117; 1.
DR PROSITE: PS00633; BROMODOMAIN_118; 1.
DR PROSITE: PS00633; BROMODOMAIN_119; 1.
DR PROSITE: PS00633; BROMODOMAIN_120; 1.
DR PROSITE: PS00633; BROMODOMAIN_121; 1.
DR PROSITE: PS00633; BROMODOMAIN_122; 1.
DR PROSITE: PS00633; BROMODOMAIN_123; 1.
DR PROSITE: PS00633; BROMODOMAIN_124; 1.
DR PROSITE: PS00633; BROMODOMAIN_125; 1.
DR PROSITE: PS00633; BROMODOMAIN_126; 1.
DR PROSITE: PS00633; BROMODOMAIN_127; 1.
DR PROSITE: PS00633; BROMODOMAIN_128; 1.
DR PROSITE: PS00633; BROMODOMAIN_129; 1.
DR PROSITE: PS00633; BROMODOMAIN_130; 1.
DR PROSITE: PS00633; BROMODOMAIN_131; 1.
DR PROSITE: PS00633; BROMODOMAIN_132; 1.
DR PROSITE: PS00633; BROMODOMAIN_133; 1.
DR PROSITE: PS00633; BROMODOMAIN_134; 1.
DR PROSITE: PS00633; BROMODOMAIN_135; 1.
DR PROSITE: PS00633; BROMODOMAIN_136; 1.
DR PROSITE: PS00633; BROMODOMAIN_137; 1.
DR PROSITE: PS00633; BROMODOMAIN_138; 1.
DR PROSITE: PS00633; BROMODOMAIN_139; 1.
DR PROSITE: PS00633; BROMODOMAIN_140; 1.
DR PROSITE: PS00633; BROMODOMAIN_141; 1.
DR PROSITE: PS00633; BROMODOMAIN_142; 1.
DR PROSITE: PS00633; BROMODOMAIN_143; 1.
DR PROSITE: PS00633; BROMODOMAIN_144; 1.
DR PROSITE: PS00633; BROMODOMAIN_145; 1.
DR PROSITE: PS00633; BROMODOMAIN_146; 1.
DR PROSITE: PS00633; BROMODOMAIN_147; 1.
DR PROSITE: PS00633; BROMODOMAIN_148; 1.
DR PROSITE: PS00633; BROMODOMAIN_149; 1.
DR PROSITE: PS00633; BROMODOMAIN_150; 1.
DR PROSITE: PS00633; BROMODOMAIN_151; 1.
DR PROSITE: PS00633; BROMODOMAIN_152; 1.
DR PROSITE: PS00633; BROMODOMAIN_153; 1.
DR PROSITE: PS00633; BROMODOMAIN_154; 1.
DR PROSITE: PS00633; BROMODOMAIN_155; 1.
DR PROSITE: PS00633; BROMODOMAIN_156; 1.
DR PROSITE: PS00633; BROMODOMAIN_157; 1.
DR PROSITE: PS00633; BROMODOMAIN_158; 1.
DR PROSITE: PS00633; BROMODOMAIN_159; 1.
DR PROSITE: PS00633; BROMODOMAIN_160; 1.
DR PROSITE: PS00633; BROMODOMAIN_161; 1.
DR PROSITE: PS00633; BROMODOMAIN_162; 1.
DR PROSITE: PS00633; BROMODOMAIN_163; 1.
DR PROSITE: PS00633; BROMODOMAIN_164; 1.
DR PROSITE: PS00633; BROMODOMAIN_165; 1.
DR PROSITE: PS00633; BROMODOMAIN_166; 1.
DR PROSITE: PS00633; BROMODOMAIN_167; 1.
DR PROSITE: PS00633; BROMODOMAIN_168; 1.
DR PROSITE: PS00633; BROMODOMAIN_169; 1.
DR PROSITE: PS00633; BROMODOMAIN_170; 1.
DR PROSITE: PS00633; BROMODOMAIN_171; 1.
DR PROSITE: PS00633; BROMODOMAIN_172; 1.
DR PROSITE: PS00633; BROMODOMAIN_173; 1.
DR PROSITE: PS00633; BROMODOMAIN_174; 1.
DR PROSITE: PS00633; BROMODOMAIN_175; 1.
DR PROSITE: PS00633; BROMODOMAIN_176; 1.
DR PROSITE: PS00633; BROMODOMAIN_177; 1.
DR PROSITE: PS00633; BROMODOMAIN_178; 1.
DR PROSITE: PS00633; BROMODOMAIN_179; 1.
DR PROSITE: PS00633; BROMODOMAIN_180; 1.
DR PROSITE: PS00633; BROMODOMAIN_181; 1.
DR PROSITE: PS00633; BROMODOMAIN_182; 1.
DR PROSITE: PS00633; BROMODOMAIN_183; 1.
DR PROSITE: PS00633; BROMODOMAIN_184; 1.
DR PROSITE: PS00633; BROMODOMAIN_185; 1.
DR PROSITE: PS00633; BROMODOMAIN_186; 1.
DR PROSITE: PS00633; BROMODOMAIN_187; 1.
DR PROSITE: PS00633; BROMODOMAIN_188; 1.
DR PROSITE: PS00633; BROMODOMAIN_189; 1.
DR PROSITE: PS00633; BROMODOMAIN_190; 1.
DR PROSITE: PS00633; BROMODOMAIN_191; 1.
DR PROSITE: PS00633; BROMODOMAIN_192; 1.
DR PROSITE: PS00633; BROMODOMAIN_193; 1.
DR PROSITE: PS00633; BROMODOMAIN_194; 1.
DR PROSITE: PS00633; BROMODOMAIN_195; 1.
DR PROSITE: PS00633; BROMODOMAIN_196; 1.
DR PROSITE: PS00633; BROMODOMAIN_197; 1.
DR PROSITE: PS00633; BROMODOMAIN_198; 1.
DR PROSITE: PS00633; BROMODOMAIN_199; 1.
DR PROSITE: PS00633; BROMODOMAIN_200; 1.
DR PROSITE: PS00633; BROMODOMAIN_201; 1.
DR PROSITE: PS00633; BROMODOMAIN_202; 1.
DR PROSITE: PS00633; BROMODOMAIN_203; 1.
DR PROSITE: PS00633; BROMODOMAIN_204; 1.
DR PROSITE: PS00633; BROMODOMAIN_205; 1.
DR PROSITE: PS00633; BROMODOMAIN_206; 1.
DR PROSITE: PS00633; BROMODOMAIN_207; 1.
DR PROSITE: PS00633; BROMODOMAIN_208; 1.
DR PROSITE: PS00633; BROMODOMAIN_209; 1.
DR PROSITE: PS00633; BROMODOMAIN_210; 1.
DR PROSITE: PS00633; BROMODOMAIN_211; 1.
DR PROSITE: PS00633; BROMODOMAIN_212; 1.
DR PROSITE: PS00633; BROMODOMAIN_213; 1.
DR PROSITE: PS00633; BROMODOMAIN_214; 1.
DR PROSITE: PS00633; BROMODOMAIN_215; 1.
DR PROSITE: PS00633; BROMODOMAIN_216; 1.
DR PROSITE: PS00633; BROMODOMAIN_217; 1.
DR PROSITE: PS00633; BROMODOMAIN_218; 1.
DR PROSITE: PS00633; BROMODOMAIN_219; 1.
DR PROSITE: PS00633; BROMODOMAIN_220; 1.
DR PROSITE: PS00633; BROMODOMAIN_221; 1.
DR PROSITE: PS00633; BROMODOMAIN_222; 1.
DR PROSITE: PS00633; BROMODOMAIN_223; 1.
DR PROSITE: PS00633; BROMODOMAIN_224; 1.
DR PROSITE: PS00633; BROMODOMAIN_225; 1.
DR PROSITE: PS00633; BROMODOMAIN_226; 1.
DR PROSITE: PS00633; BROMODOMAIN_227; 1.
DR PROSITE: PS00633; BROMODOMAIN_228; 1.
DR PROSITE: PS00633; BROMODOMAIN_229; 1.
DR PROSITE: PS00633; BROMODOMAIN_230; 1.
DR PROSITE: PS00633; BROMODOMAIN_231; 1.
DR PROSITE: PS00633; BROMODOMAIN_232; 1.
DR PROSITE: PS00633; BROMODOMAIN_233; 1.
DR PROSITE: PS00633; BROMODOMAIN_234; 1.
DR PROSITE: PS00633; BROMODOMAIN_235; 1.
DR PROSITE: PS00633; BROMODOMAIN_236; 1.
DR PROSITE: PS00633; BROMODOMAIN_237; 1.
DR PROSITE: PS00633; BROMODOMAIN_238; 1.
DR PROSITE: PS00633; BROMODOMAIN_239; 1.
DR PROSITE: PS00633; BROMODOMAIN_240; 1.
DR PROSITE: PS00633; BROMODOMAIN_241; 1.
DR PROSITE: PS00633; BROMODOMAIN_242; 1.
DR PROSITE: PS00633; BROMODOMAIN_243; 1.
DR PROSITE: PS00633; BROMODOMAIN_244; 1.
DR PROSITE: PS00633; BROMODOMAIN_245; 1.
DR PROSITE: PS00633; BROMODOMAIN_246; 1.
DR PROSITE: PS00633; BROMODOMAIN_247; 1.
DR PROSITE: PS00633; BROMODOMAIN_248; 1.
DR PROSITE: PS00633; BROMODOMAIN_249; 1.
DR PROSITE: PS00633; BROMODOMAIN_250; 1.
DR PROSITE: PS00633; BROMODOMAIN_251; 1.
DR PROSITE: PS00633; BROMODOMAIN_252; 1.
DR PROSITE: PS00633; BROMODOMAIN_253; 1.
DR PROSITE: PS00633; BROMODOMAIN_254; 1.
DR PROSITE: PS00633; BROMODOMAIN_255; 1.
DR PROSITE: PS00633; BROMODOMAIN_256; 1.
DR PROSITE: PS00633; BROMODOMAIN_257; 1.
DR PROSITE: PS00633; BROMODOMAIN_258; 1.
DR PROSITE: PS00633; BROMODOMAIN_259; 1.
DR PROSITE: PS00633; BROMODOMAIN_260; 1.
DR PROSITE: PS00633; BROMODOMAIN_261; 1.
DR PROSITE: PS00633; BROMODOMAIN_262; 1.
DR PROSITE: PS00633; BROMODOMAIN_263; 1.
DR PROSITE: PS00633; BROMODOMAIN_264; 1.
DR PROSITE: PS00633; BROMODOMAIN_265; 1.
DR PROSITE: PS00633; BROMODOMAIN_266; 1.
DR PROSITE: PS00633; BROMODOMAIN_267; 1.
DR PROSITE: PS00633; BROMODOMAIN_268; 1.
DR PROSITE: PS00633; BROMODOMAIN_269; 1.
DR PROSITE: PS00633; BROMODOMAIN_270; 1.
DR PROSITE: PS00633; BROMODOMAIN_271; 1.
DR PROSITE: PS00633; BROMODOMAIN_272; 1.
DR PROSITE: PS00633; BROMODOMAIN_273; 1.
DR PROSITE: PS00633; BROMODOMAIN_274; 1.
DR PROSITE: PS00633; BROMODOMAIN_275; 1.
DR PROSITE: PS00633; BROMODOMAIN_276; 1.
DR PROSITE: PS00633; BROMODOMAIN_277; 1.
DR PROSITE: PS00633; BROMODOMAIN_278; 1.
DR PROSITE: PS00633; BROMODOMAIN_279; 1.
DR PROSITE: PS00633; BROMODOMAIN_280; 1.
DR PROSITE: PS00633; BROMODOMAIN_281; 1.
DR PROSITE: PS00633; BROMODOMAIN_282; 1.
DR PROSITE: PS00633; BROMODOMAIN_283; 1.
DR PROSITE: PS00633; BROMODOMAIN_284; 1.
DR PROSITE: PS00633; BROMODOMAIN_285; 1.
DR PROSITE: PS00633; BROMODOMAIN_286; 1.
DR PROSITE: PS00633; BROMODOMAIN_287; 1.
DR PROSITE: PS00633; BROMODOMAIN_288; 1.
DR PROSITE: PS00633; BROMODOMAIN_289; 1.
DR PROSITE: PS00633; BROMODOMAIN_290; 1.
DR PROSITE: PS00633; BROMODOMAIN_291; 1.
DR PROSITE: PS00633; BROMODOMAIN_292; 1.
DR PROSITE: PS00633; BROMODOMAIN_293; 1.
DR PROSITE: PS00633; BROMODOMAIN_294; 1.
DR PROSITE: PS00633; BROMODOMAIN_295; 1.
DR PROSITE: PS00633; BROMODOMAIN_296; 1.
DR PROSITE: PS00633; BROMODOMAIN_297; 1.
DR PROSITE: PS00633; BROMODOMAIN_298; 1.
DR PROSITE: PS00633; BROMODOMAIN_299; 1.
DR PROSITE: PS00633; BROMODOMAIN_300; 1.
DR PROSITE: PS00633; BROMODOMAIN_301; 1.
DR PROSITE: PS00633; BROMODOMAIN_302; 1.
DR PROSITE: PS00633; BROMODOMAIN_303; 1.
DR PROSITE: PS00633; BROMODOMAIN_304; 1.
DR PROSITE: PS00633; BROMODOMAIN_305; 1.
DR PROSITE: PS00633; BROMODOMAIN_306; 1.
DR PROSITE: PS00633; BROMODOMAIN_307; 1.
DR PROSITE: PS00633; BROMODOMAIN_308; 1.
DR PROSITE: PS00633; BROMODOMAIN_309; 1.
DR PROSITE: PS00633; BROMODOMAIN_310; 1.
DR PROSITE: PS00633; BROMODOMAIN_311; 1.
DR PROSITE: PS00633; BROMODOMAIN_312; 1.
DR PROSITE: PS00633; BROMODOMAIN_313; 1.
DR PROSITE: PS00633; BROMODOMAIN_314; 1.
DR PROSITE: PS00633; BROMODOMAIN_315; 1.
DR PROSITE: PS00633; BROMODOMAIN_316; 1.
DR PROSITE: PS00633; BROMODOMAIN_317; 1.
DR PROSITE: PS00633; BROMODOMAIN_318; 1.
DR PROSITE: PS00633; BROMODOMAIN_319; 1.
DR PROSITE: PS00633; BROMODOMAIN_320; 1.
DR PROSITE: PS00633; BROMODOMAIN_321; 1.
DR PROSITE: PS00633; BROMODOMAIN_322; 1.
DR PROSITE: PS00633; BROMODOMAIN_323; 1.
DR PROSITE: PS00633; BROMODOMAIN_324; 1.
DR PROSITE: PS00633; BROMODOMAIN_325; 1.
DR PROSITE: PS00633; BROMODOMAIN_326; 1.
DR PROSITE: PS00633; BROMODOMAIN_327; 1.
DR PROSITE: PS00633; BROMODOMAIN_328; 1.
DR PROSITE: PS00633; BROMODOMAIN_329; 1.
DR PROSITE: PS00633; BROMODOMAIN_330; 1.
DR PROSITE: PS00633; BROMODOMAIN_331; 1.
DR PROSITE: PS00633; BROMODOMAIN_332; 1.
DR PROSITE: PS00633; BROMODOMAIN_333; 1.
DR PROSITE: PS00633; BROMODOMAIN_334; 1.
DR PROSITE: PS00633; BROMODOMAIN_335; 1.
DR PROSITE: PS00633; BROMODOMAIN_336; 1.
DR PROSITE: PS00633; BROMODOMAIN_337; 1.
DR PROSITE: PS00633; BROMODOMAIN_338; 1.
DR PROSITE: PS00633; BROMODOMAIN_339; 1.
DR PROSITE: PS00633; BROMODOMAIN_340; 1.
DR PROSITE: PS00633; BROMODOMAIN_341; 1.
DR PROSITE: PS00633; BROMODOMAIN_342; 1.
DR PROSITE: PS00633; BROMODOMAIN_343; 1.
DR PROSITE: PS00633; BROMODOMAIN_344; 1.
DR PROSITE: PS00633; BROMODOMAIN_345; 1.
DR PROSITE: PS00633; BROMODOMAIN_346; 1.
DR PROSITE: PS00633; BROMODOMAIN_347; 1.
DR PROSITE: PS00633; BROMODOMAIN_348; 1.
DR PROSITE: PS00633; BROMODOMAIN_349; 1.
DR PROSITE: PS00633; BROMODOMAIN_350; 1.
DR PROSITE: PS00633; BROMODOMAIN_351; 1.
DR PROSITE: PS00633; BROMODOMAIN_352; 1.
DR PROSITE: PS00633; BROMODOMAIN_353; 1.
DR PROSITE: PS00633; BROMODOMAIN_354; 1.
DR PROSITE: PS00633; BROMODOMAIN_355; 1.
DR PROSITE: PS00633; BROMODOMAIN_356; 1.
DR PROSITE: PS00633; BROMODOMAIN_357; 1.
DR PROSITE: PS00633; BROMODOMAIN_358; 1.
DR PROSITE: PS00633; BROMODOMAIN_359; 1.
DR PROSITE: PS00633; BROMODOMAIN_360; 1.
DR PROSITE: PS00633; BROMODOMAIN_361; 1.
DR PROSITE: PS00633; BROMODOMAIN_362; 1.
DR PROSITE: PS00633; BROMODOMAIN_363; 1.
DR PROSITE: PS00633; BROMODOMAIN_364; 1.
DR PROSITE: PS00633; BROMODOMAIN_365; 1.
DR PROSITE: PS00633; BROMODOMAIN_366; 1.
DR PROSITE: PS00633; BROMODOMAIN_367; 1.
DR PROSITE: PS00633; BROMODOMAIN_368; 1.
DR PROSITE: PS00633; BROMODOMAIN_369; 1.
DR PROSITE: PS00633; BROMODOMAIN_370; 1.
DR PROSITE: PS00633; BROMODOMAIN_371; 1.
DR PROSITE: PS00633; BROMODOMAIN_372; 1.
DR PROSITE: PS00633; BROMODOMAIN_373; 1.
DR PROSITE: PS00633; BROMODOMAIN_374; 1.
DR PROSITE: PS00633; BROMODOMAIN_375; 1.
DR PROSITE: PS00633; BROMODOMAIN_376; 1.
DR PROSITE: PS00633; BROMODOMAIN_377; 1.
DR PROSITE: PS00633; BROMODOMAIN_378; 1.
DR PROSITE: PS00633; BROMODOMAIN_379; 1.
DR PROSITE: PS00633; BROMODOMAIN_380; 1.
DR PROSITE: PS00633; BROMODOMAIN_381; 1.
DR PROSITE: PS00633; BROMODOMAIN_382; 1.
DR PROSITE: PS00633; BROMODOMAIN_383; 1.
DR PROSITE: PS00633; BROMODOMAIN_384; 1.
DR PROSITE: PS00633; BROMODOMAIN_385; 1.
DR PROSITE: PS00633; BROMODOMAIN_386; 1.
DR PROSITE: PS00633; BROMODOMAIN_387; 1.
DR PROSITE: PS00633; BROMODOMAIN_388; 1.
DR PROSITE: PS00633; BROMODOMAIN_389; 1.
DR PROSITE: PS00633; BROMODOMAIN_390; 1.
DR PROSITE: PS00633; BROMODOMAIN_391; 1.
DR PROSITE: PS00633; BROMODOMAIN_392; 1.
DR PROSITE: PS00633; BROMODOMAIN_393; 1.
DR PROSITE: PS00633; BROMODOMAIN_394; 1.
DR PROSITE: PS00633; BROMODOMAIN_395; 1.
DR PROSITE: PS00633; BROMODOMAIN_396; 1.
DR PROSITE: PS00633; BROMODOMAIN_397; 1.
DR PROSITE: PS00633; BROMODOMAIN_398; 1.
DR PROSITE: PS00633; BROMODOMAIN_399; 1.
DR PROSITE: PS00633; BROMODOMAIN_400; 1.
DR PROSITE: PS00633; BROMODOMAIN_401; 1.
DR PROSITE: PS00633; BROMODOMAIN_402; 1.
DR PROSITE: PS00633; BROMODOMAIN_403; 1.
DR PROSITE: PS00633; BROMODOMAIN_404; 1.
DR PROSITE: PS00633; BROMODOMAIN_405; 1.
DR PROSITE: PS00633; BROMODOMAIN_406; 1.
DR PROSITE: PS00633; BROMODOMAIN_407; 1.
DR PROSITE: PS00633; BROMODOMAIN_408; 1.
DR PROSITE: PS00633; BROMODOMAIN_409; 1.
DR PROSITE: PS00633; BROMODOMAIN_410

DE P54 protein precursor.
 OS Enterococcus faecium (Streptococcus faecium).
 OC Bacteria; Firmicutes; Lactobacillales; Enterococcaceae; Enterococcus.
 OX NCBI_TaxID=1352;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=89385998; PubMed=2780297;
 RX Fuerst P., Moesch H.-U., Solioz M.;
 RT "A protein of unusual composition from Enterococcus faecium";
 RL Nucleic Acids Res. 17:6724-6724(1989).
 CC -1- SUBCELLULAR LOCATION: Attached to the cell wall.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY C40.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, X16421; CAA34442.1; ALT_INIT.
 CC DR PIR, S05542; S05542.
 CC DR MEROPS; C40.UFW;
 CC DR InterPro; IPR000064; NLP_C_P60.
 CC DR Pfam; PF00877; NLP_C_P60; 1.
 CC KW Signal; Cell wall.
 CC FT SIGNAL 1 27 POTENTIAL.
 FT CHAIN 28 516 P54 PROTEIN.
 FT ACT SITE 429 429 POTENTIAL.
 SQ SEQUENCE 516 AA; 54596 MW; 402ECAA439846D26 CRC64;
 Query Match 3.4%; Score 140.5; DB 1; Length 516;
 Best Local Similarity 19.0%; Pred. No. 1.5; Index 153; Gaps 17;
 Matches 88; Conservative 70; Mismatches 151;
 QY 8 QAR-TYKENNRVSYIDGKQATKTNLTPEVSKREG-INAEOIVIKITDQGVYSHGDH 65
 DB 107 QARETVKNTSSNYIDAV-----LNMDSLADANGRIQAMSTYIK-ANQDLVQOQKD 157
 QY 66 YHYVNGKVYYDAIIEELMKDNYLKKEDIVNEVKGVIYK-----DGKYV 115
 DB 158 KOAVEAKKANEKOKELADNQALLESOKGDLAKQADLVLTSLAEQATADKADL 217
 QY 116 YLKDA-AHADNVRTKE---INRQKHSHRGCGTPRDGAVALASGGRYTTDDGIYF 171
 DB 218 NRKKAFAEAEOAIRQOARLAEQARQOAAQKAEKAREQOAAQAAQTA-----LS 269
 QY 172 NASDIIEDGDAVIVHGDIHYIIPNELSASELAFAFLSGRNLNSRTYRONSDN 231
 DB 270 SASITTESSAA-----QSSSESKAPESSTTTEESTSTSTSTTSSSTG 314
 QY 232 TSRT-----NMVP-----SVNPGTTNTNTNNSNTNSQASQNDIDSLKQ 273
 DB 315 SSSSTESSSTESTVPESSSTOESTPANTESSSSSSNTNNNTNNNTNNSTG 370
 QY 274 LYKLPLSQHVESDGLVFPDQITSTAGVAVPHGDHFIPIYSQMSLEERIAIIFL 333
 DB 371 ----- 370
 QY 334 RYRSHHWVDSRPEQSPQPPSPGPAPALIKIDSN-SSIVSOLVAKVGEVY----- 388
 DB 371 NNNNNNTV-----TPAPTPPTPAPAPAPRPPSSVGAIVAAVYKIGIPYWGK 422
 QY 389 ---FEKGISRYV---AKDLPSFTV--KNLESLSKQES 420
 DB 423 DSGGFCGSGFTRYVYLQVTRDIDGWTVPQESAGTKISVSOA 464

DT 01-JUL-1989 (Rel. 11, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Laminin gamma-1 chain precursor (laminin B2 chain).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=91093128; PubMed=1985895;
 RX Kallunki T., Ikonen J., Chow L.T., Kallunki P., Trygvaason K.;
 RT "Structure of the human laminin B2 chain gene reveals extensive
 RT divergence from the laminin B1 chain gene";
 RL J. Biol. Chem. 266:221-228(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=88198245; PubMed=3360804;
 RX Pikkariainen T., Kallunki T., Trygvaason K.;
 RT "Human laminin B2 chain. Comparison of the complete amino acid
 RT sequence with the B1 chain reveals variability in sequence homology
 RT between different structural domains";
 RL J. Biol. Chem. 263:6751-6758(1988).
 RN [3]
 RP SEQUENCE OF 1393-1609 FROM N.A.
 RX MEDLINE=89169663; PubMed=3234037;
 RA Fukushima Y., Pikkariainen T., Kallunki T., Eddy R.L., Byers M.G.,
 RA Haley L.L., Henry W.M., Trygvaason K., Shows T.B.;
 RT "Isolation of a human laminin B2 (LAMB2) cDNA clone and assignment of
 RT the gene to chromosome region 1q25--q31.";
 RL Cytogenet. Cell Genet. 48:137-141(1988).
 RN [4]
 RP SEQUENCE OF 1282-1609 FROM N.A.
 RC TISSUE=Endothelial cells;
 RX MEDLINE=92216129; PubMed=1806043;
 RA Santos C.L.S., Sabaga J., Brentani R.;
 RT "Differences in human laminin B2 sequences.";
 RL DNA Seq. 1:275-277(1991).
 CC -1- FUNCTION: Binding to cells via a high affinity receptor, laminin
 CC is thought to mediate the attachment, migration, and organization
 CC of cells into tissues during embryonic development by interacting
 CC with other extracellular matrix components.
 CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
 CC different polypeptide chains (alpha, beta, gamma), which are bound
 CC to each other by disulfide bonds into a cross-shaped molecule
 CC comprising one long and three short arms with globules at each
 CC end.
 CC THE GAMMA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (BHS LAMININ),
 CC LAMININ-2 (MEROSIN), LAMININ-3 (S-LAMININ), LAMININ-4 (S-MEROSIN),
 CC LAMININ-6 (K-LAMININ) AND LAMININ-7 (KS-LAMININ).
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
 CC COMPONENT).
 CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
 CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
 CC -1- SIMILARITY: Contains 1 laminin N-terminal domain.
 CC -1- SIMILARITY: Contains 11 laminin EGF-like domains.
 CC -1- SIMILARITY: Contains 1 laminin IV domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL, M55210; AAA59492.1;
 DR EMBL, M55217; AAA59492.1; JOINED.
 DR EMBL, M55201; AAA59492.1; JOINED.
 DR EMBL, M55211; AAA59492.1; JOINED.

DR	EMBL	MS5212	AAAS9492.1	JOINED	
DR	EMBL	MS5213	AAAS9492.1	JOINED	
DR	EMBL	MS5214	AAAS9492.1	JOINED	
DR	EMBL	MS5215	AAAS9492.1	JOINED	
DR	EMBL	MS5216	AAAS9492.1	JOINED	
DR	EMBL	MS5192	AAAS9492.1	JOINED	
DR	EMBL	MS5193	AAAS9492.1	JOINED	
DR	EMBL	MS5194	AAAS9492.1	JOINED	
DR	EMBL	MS5195	AAAS9492.1	JOINED	
DR	EMBL	MS5196	AAAS9492.1	JOINED	
DR	EMBL	MS5197	AAAS9492.1	JOINED	
DR	EMBL	MS5198	AAAS9492.1	JOINED	
DR	EMBL	MS5199	AAAS9492.1	JOINED	
DR	EMBL	MS5200	AAAS9492.1	JOINED	
DR	EMBL	MS5202	AAAS9492.1	JOINED	
DR	EMBL	MS5203	AAAS9492.1	JOINED	
DR	EMBL	MS5204	AAAS9492.1	JOINED	
DR	EMBL	MS5205	AAAS9492.1	JOINED	
DR	EMBL	MS5206	AAAS9492.1	JOINED	
DR	EMBL	MS5207	AAAS9492.1	JOINED	
DR	EMBL	MS5208	AAAS9492.1	JOINED	
DR	EMBL	MS5209	AAAS9492.1	JOINED	
DR	EMBL	J03302	AAAS9488.1	-	
DR	EMBL	M27654	AAAS9489.1	-	
DR	EMBL	X13939	CAA32122.1	-	
DR	PIR	S13548	MMHUB2		
DR	HSSP	P02468	ITLE		
DR	GIMM	HGNC:6492	LAMC1		
DR	GO	GO:0005644	C:basement membrane; TAS.		
DR	GO	GO:0007492	P:endothelium development; TAS.		
DR	GO	GO:0006461	P:protein complex assembly; TAS.		
DR	InterPro	IPR006209	EGF-like		
DR	InterPro	IPR000034	Laminin_B		
DR	InterPro	IPR002049	Laminin_EGF		
DR	InterPro	IPR001866	LamNT		
DR	pfam	PF00052	Laminin_B_1		
DR	pfam	PF00053	Laminin_EGF_10		
DR	pfam	PF00055	Laminin_Nterm_1		
DR	ProDom	PD002082	Lam_N2_1		
DR	SMART	SM00180	EGF_Lam_8		
DR	SMART	SM00281	LamB_1		
DR	SMART	SM00136	LamNT_1		
DR	PROSITE	PS00022	EGF_1_8		
DR	PROSITE	PS01186	EGF_2_2		
DR	PROSITE	PS01248	LAMININ TYPE EGF_11		
KW	Glycoprotein	Basement membrane; Extracellular matrix; Coiled coil; Laminin EGF-like domain; Cell adhesion; Repeat; Signal; Polymorphism			
KW	Laminin EGF-like domain				
FT	SIGNAL	1	33		
FT	CHAIN	34	1609	LAMININ GAMMA-1 CHAIN.	
FT	DOMAIN	34	285	LAMININ N-TERMINAL (DOMAIN VII)	
FT	DOMAIN	286	341	LAMININ EGF-LIKE 1.	
FT	DOMAIN	342	397	LAMININ EGF-LIKE 2.	
FT	DOMAIN	398	444	LAMININ EGF-LIKE 3.	
FT	DOMAIN	445	494	LAMININ EGF-LIKE 4.	
FT	DOMAIN	495	504	LAMININ EGF-LIKE 5 (N-TERMINAL)	
FT	DOMAIN	505	689	LAMININ EGF-LIKE 6	
FT	DOMAIN	690	723	LAMININ EGF-LIKE 5 (C-TERMINAL)	
FT	DOMAIN	724	772	LAMININ EGF-LIKE 6	
FT	DOMAIN	773	827	LAMININ EGF-LIKE 7	
FT	DOMAIN	828	883	LAMININ EGF-LIKE 8	
FT	DOMAIN	884	934	LAMININ EGF-LIKE 9	
FT	DOMAIN	935	982	LAMININ EGF-LIKE 10	
FT	DOMAIN	983	1030	LAMININ EGF-LIKE 11	
FT	DOMAIN	1030	1609	DOMAIN II (AND I)	
FT	DOMAIN	1038	1609	COILED COIL (POTENTIAL)	
FT	DISULFID	286	295	BY SIMILARITY.	
FT	DISULFID	288	305	BY SIMILARITY.	
FT	DISULFID	307	316	BY SIMILARITY.	
FT	DISULFID	319	339	BY SIMILARITY.	
FT	DISULFID	342	351	BY SIMILARITY.	
FT	DISULFID	344	367	BY SIMILARITY.	
FT	DISULFID	344	379	BY SIMILARITY.	

Query Match	Similarity	Score	DB 1	Length
Best Local	20.1%	139.5		1609
Matches	98	Pred. No. 8.1		
	Conservative	70	Mismatches	194
			Indels	125
			Gaps	19
359	PGCPAP-----	-----NLKIDNSSLSVSQLVRKVGEGVVEBKISRYVFAKDLP	35	405
1026	PGCECEPCRYRVLVKQVADHRRVKQLELSIANL--GTGDEM-----	-----	1066	
406	ETVNLKESKLSKQESVSHTLTAKKENVAAPDOEFYDKAVNLTJAHKALFYNGKGRNSDFQ	465		
1067	-TDAFEDRLKEAREVWMDLREAOVDVQNMMDRQGRNNLTLSQISRLQNRITIE	1122		
466	ALDYLLERLNDGSTNKEKLV-----DLAFLAPITHPRLGKNSOI-----	508		
1126	ETGNILAEQARAHVNTERRLIEIASRELEKAKVAANVSVTPESTGDPNNMTLLAEBAEK	1185		
509	-----ETDEBVAIQAQADKYTTSDDGTFPEHNDIISBQDAYVTPRHGHSHWIKSDLS	562		
1186	LAERHKQEDADIVVAKTAN--TSTEAANNLLRTLAGNQTAFIEELNRYKQAKNSIQ	1244		

```

QY 563 DEXKAAQAQYTKKGIPLPSPADVYKANTGDSAAIYVRKGEKGIPLVRLPYMVEHTV 622
DB 1245 DLEKQAAARVHE-----EAKRAGDAKVAIVASV-----AQLSPIDSETL 1282
QY 623 EYKNGNL-----IIPHK-DHYHNIFAMFDHTYK-----APNGYTLLEL 661
DB 1283 EHEANNIKKEAENLEOLIOCKLKVYDLR---EDRGRGELEVKNLLEKGTQOQADOL 1338
QY 662 FA---TIKYVEHPDERPHSNDGNGNASEHVLGKKHSDPNNKFADEPVEETPAEPE 718
DB 1339 IARADAARALAEBAAK--GRDTLOEANDILNLIKPFDRRVNDKTRAAAEALRKIPA---- 1393
QY 719 VPQVETEKVEAOLKEAEVLL---AKYVDSLKANAETFLA-GLRNNLTLOIMNNSIMA 773
DB 1394 INQITTEANE-KTREAOQALGSAADATKAKNAKHAERIASAVQGNAT-----STVA 1445
QY 774 EAEKLLA 780
DB 1446 EAERTFA 1452

RESULT 30
SPOF_SCHPO STANDARD: PRT; 1957 AA.
AC Q10411; Q9USE9;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Sporulation-specific protein 15.
GN SP015 OR SPAC1F3.06C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxId=4896;

RN (1)
RP SEQUENCE FROM N.A., FUNCTION, SUBUNIT, AND SUBCELLULAR LOCATION.
RX MEDLINE=20107136; PubMed=10639340;
RA Ikemoto S., Nakamura T., Kubo M., Shimoda C.;
RT "S. pombe sporulation-specific coiled-coil protein Spo15p is localized
RL to the spindle pole body and essential for its modification."
RL J. Cell Sci. 113:545-554(2000).
RN (2)
RP SEQUENCE FROM N.A.
RX STRAIN=972;
RC MEDLINE=21848401; PubMed=11859360;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgouros J., Peat N., Hayles J., Baker S., Baaham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney P., Moutre S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Alvert R., Robben J., Grymoprepz B.,
RA Welljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,
RA Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Medler H., Wandut R., Purnelle B.,
RA Goffeau A., Cadieu E., Drenth S., Gloux S., Lelaure V., Mottier S.,
RA Gallbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;
RT "The genome sequence of Schizosaccharomyces pombe."
RL Nature 415:871-880(2002).
RN (3)

```

```

RP SEQUENCE OF 705-871 FROM N.A.
RC STRAIN=968 h90;
RX MEDLINE=20223868; PubMed=10759889;
RA Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Harauchi T.,
RA Hirooka Y.;
RT "Large-scale screening of intracellular protein localization in living
RT fission yeast cells by the use of a GFP-fusion genomic DNA library."
RT Genes Cells 5:169-190(2000).
CC -1- FUNCTION: Has a role in the initiation of spore membrane
CC formation.
CC -1- SUBUNIT: Monomer.
CC -1- SUBCELLULAR LOCATION: Spindle pole body.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z70690; CA94624.1; -
CC EMBL: AB027811; BA87115.1; -.
CC PIR: T38077; T38077.
CC GeneDB: SPombe; SPAC1F3.06c; -.
CC Sporulation; Coiled coil.
CC DOMAIN 199 785 COILED COIL (POTENTIAL).
CC FT DOMAIN 804 1235 COILED COIL (POTENTIAL).
CC FT DOMAIN 1320 1471 COILED COIL (POTENTIAL).
CC FT DOMAIN 1481 1723 COILED COIL (POTENTIAL).
CC SQ SEQUENCE 1957 AA; 222785 MW; 3F480CA06171BD9A CRC64;

```

```

Query Match 3.3%; Score 139; DB 1; Length 1957;
Best local similarity 18.2%; Pred. No. 11;
Matches 159; Conservative 143; Mismatches 330; Indels 244; Gaps 36;

QY 81 EELLKMDPYKLDKEDIVNEVKG--GYVYKVGKYYVYLKDAAHAD---NVRTKEEINR 134
DB 640 QQLQLEBRANFEQKESTLSDENNDLRTKLKLEESKSLIKQEDVDSLEKNIQTLEKDR 699
QY 135 QKQHSQHBEGTTPNDGAVNALSQGRYTTDDGYTFNNSDIIETDGYVYPHGDHYHY 194
DB 700 KSEE-----ALRFSKLEAK-----MLREVIDML-----KQKHETL 729
QY 195 -IPKREL--SASELAAEAFLSGRGNLSN-----SRYRRONSNTSRTMNV 238
DB 730 EAQRNDLHSSLSDAKNTNALISSELTKESEDEVKRLTANVETLTQDSKAMKQSFSLVNSY 789
QY 239 PSVSN-----PGTNTNTSNNSNTNSQASQNDISLKLQLYKL-----PSQRHVESD 287
DB 790 QSISLNLHRLRDRDHNMOSQNTLLESSEKLTDCENTLQCMNTLIDNVQKLMKHVNOE 849
QY 288 GLVFPDAQITS-----FTARGVAVPHGDHYHYFIPYSQNSELEERLARIIPRYRSN 338
DB 850 SKVSELEKVENGLSLDLKRLRSLSLVVAISDNDQI---LTQALSELSKVYDSLEQSAQLN 905
QY 339 HWVPDSRREPQSPQPTPEPSPQAPNLIKID-----SNSLSVLQVRKVGEGY 387
DB 906 SGLKSLAEAKOHLATENEL-----HIRLDKLTGKLKIEESKSSDLKTLTAQOEIS 958
QY 388 VEEKGISRYVPAKDLPSRTVNLKESLKSQESVHTLTAKKENAVAPRQEFYDAVNL 447
DB 950 NIKENMSQ-----SQATISVSKLDETLSSKSLKLEADLEHLKAKKSEVEVERNALL 1010
QY 448 TEAHKALFKNKGRNSDFALDKLLERL--NDESTNKEKLYD--DLAFLAPITHPERL 501
DB 1011 ASNERLMDLKNNGENIASLQTEIRKRAENDDLQSKLSVSVSEYENLLISSQNKNSL 1070
QY 502 GKPNOIEYTE-----DE-----VRIAQLADKTTSDGYTFPDEHDIISDGDVATPH 549
DB 1071 DKTN-QLKYTEKNVQKLDEKQORVVELEELSKY-----GKLGEENAOIKELLAL-LRKK 1124
QY 550 MGHSHWIGKDSLSD-KEKV-AAQAATKEKGIPLPS----- 582

```

```

Db      1125 SKQDLCANFVDDLEKESDALEQLTNEKNELIVLEQSNNSNEALVEERDLANRLSDM 1184
Oy      503 ----PDAD-----VKAN-----PTGDSAAAIYNRV-----603
Db      1185 KKSJSDSDNVIASIRSDLVARNVDELDTLKQKDSISTOYSEVCGQDRDDLDSLKGCESBF 1244
Oy      604 -----KGEKRIPLVRLPYVMEHTVEVKGNL-----IIPHGDYHNHK 641
Db      1245 NKVAVSLRELCTKSIDVPVSEI---LDDNPFNAGNSELSRLTVLENTLDFNQVN 1301
Oy      642 FAWPDDHTYKAPNGYTLDELFPATIKYVVEHPDERPHSNDGW---GNASEHVLGKQHS 697
Db      1302 F-----KKMELDNRLTTTD--AEFTKVVADELEKLQHEHDDWMLIQRGDLEKAL-----K 1347
Oy      698 DPNKRFKADDERVEETPAPEPEVQVETEKVEKQLEAEVLAKYDSSUKANATTTAGL 757
Db      1348 DSEKNFLRKEKEMENINISLEEGKEETKEIAELIS-----SRLEDNOL---ATNKLKNQ 1398
Oy      758 RNNLTLOIMDNNSIMAAEAKLTALKGSNPSSVSKE 793
Db      1399 LDHLNQEIRLKEDEVLEKESLILISLEESLSNORQKE 1434

RESULT 31
SNF2_YEAST STANDARD: PRT: 1703 AA.
AC P22082,
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Transcription regulatory protein SNF2 (SWI/SNF complex component SNF2)
DE (Regulatory protein SWI2) (Regulatory protein GAML1) (Transcription
DE factor TYB3).
GN SNF2 OR SWI2 OR GAML1 OR TYB3 OR RIC1 OR YOR290C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OC NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=91187857; PubMed=1901413;
RA Laurent B.C., Treitel M.A., Carlson M.;
RT "Functional interdependence of the yeast SNF2, SNF5, and SNF6
RT proteins in transcriptional activation.";
RT Proc. Natl. Acad. Sci. U.S.A. 88:2687-2691(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AH22;
RX MEDLINE=91360076; PubMed=186612;
RA Yoshimoto H., Yamashita I.;
RT "The GAML1/SNF2 gene of Saccharomyces cerevisiae encodes a highly
RT charged nuclear protein required for transcription of the STA1
RT gene.";
RT Mol. Gen. Genet. 228:270-280(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1B;
RX MEDLINE=95332261; PubMed=7608126;
RA Kodaki T., Hosaka K., Nakawa J., Yamashita S.;
RT "The SNF2/SWI2/GAML1/TYB3/RIC1 gene is involved in the coordinate
RT regulation of phospholipid synthesis in Saccharomyces cerevisiae.";
RT J. Biochem. 117:362-368(1995).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=97051594; PubMed=8896271;
RA Cheret G., Bernardi A., Sor F.J.;
RT "DNA sequence analysis of the VP1-SNF2 region on chromosome XV of
RT Saccharomyces cerevisiae.";
RT Yeast 12:1059-1064(1996).
RN [5]

```

```

RP SEQUENCE OF 1-309 FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=97298310; PubMed=9153758;
RA Poirey R., Cziepluch C., Tobiasch E., Pujol A., Kordes E.,
RA Jauniaux J.-C.;
RT "Sequence and analysis of a 36.2 kb fragment from the right arm of
RT yeast chromosome XV reveals 19 open reading frames including SNF2 (5'
RT end), CPA1, SLY41, a putative transport ATPase, a putative ribosomal
RT protein and an SNF2 homologue.";
RL Yeast 13:479-482(1997).
RN [6]
RP MUTAGENESIS.
RX MEDLINE=97025355; PubMed=8871545;
RA Richmond E., Peterson C.L.;
RT "Functional analysis of the DNA-stimulated ATPase domain of yeast
RT SWI2/SNF2.";
RL Nucleic Acids Res. 24:3685-3692(1996).
CC - FUNCTION: INVOLVED IN TRANSCRIPTIONAL ACTIVATION. THE SWI/SNF
CC COMPLEX IS REQUIRED FOR THE INDUCED EXPRESSION OF A LARGE NUMBER
CC OF GENES. THIS COMPLEX ALTERS CHROMATIN STRUCTURE TO FACILITATE
CC BINDING OF GENE-SPECIFIC DEDICATED TRANSCRIPTION FACTORS.
CC - SUBUNIT: COMPONENT OF THE SWI/SNF GLOBAL TRANSCRIPTION ACTIVATOR
CC COMPLEX.
CC - SUBCELLULAR LOCATION: Nuclear.
CC - SIMILARITY: STRONG, TO DROSOPHILA BRAHMA.
CC - SIMILARITY: Contains 1 bromodomain.
CC - SIMILARITY: Contains 3 A.T hook DNA-binding repeats.
CC - SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M61703; AAA5059.1; -
DR EMBL; X57837; CAA40969.1; -
DR EMBL; D90459; BAA14423.1; -
DR EMBL; X89633; CAA61793.1; -
DR EMBL; Z75198; CAA95917.1; -
DR EMBL; Z75199; CAA95919.1; -
DR PIR; S15047; S15047.
DR TRANSFAC; T02401; -.
DR SGD; S0005816; SNF2.
DR InterPro; IPR000637; AT_hook.
DR InterPro; IPR001487; Bromodomain.
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF02178; AT_hook_2.
DR Pfam; PF00439; bromodomain_1.
DR Pfam; PF00271; helicase_C_1.
DR Pfam; PF00176; SNF2_N_1.
DR PRINTS; PR00929; ATHOOK.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00384; AT_hook_2.
DR SMART; SM00297; BROMO_1.
DR SMART; SM00487; DEXDC_1.
DR SMART; SM00490; HELIC_C_1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
KM Transcription regulation; Nuclear protein; Activator; Repeat;
KW Bromodomain; ATP-binding; Helicase.
FT DOMAIN 55 68 GLN-RICH.
FT DOMAIN 207 239 ALA/GLN-RICH.
FT NP_BIND 792 799 ATP (BY SIMILARITY).
FT SITE 894 897 DEGH BOX.
FT DNA_BIND 1446 1456 A.T HOOK 1.
FT DNA_BIND 1502 1513 A.T HOOK 2.
FT DNA_BIND 1516 1526 A.T HOOK 3.
FT DOMAIN 1568 1638 BROMODOMAIN.

```



```

Db 220 LVFGYRKPLDNIKDNVGMEDYIKKKK-----KTIENINELIEBSKTTIKN 266
Qy 419 ESVSHTLAKKENVAPROEFDK-----AYNLLTEAHKALFXNKNRNSDFQALDILRL 474
Db 267 KNATIEEEKKLYQAQYDLSIYNKQLEENHLISLEKRIIDLKKNENIKELLDKINEIK 326
Qy 475 NDESTNKEKLYVDLLAFAPLIPTRLGKPNQSIETDEVRIAQADKYTTSDGYIPDE 534
Db 327 NPPRANSNT-----PNTLLDKKKKIEHEHEKEIKTKIKFN----- 365
Qy 535 HDIISDEGDVAYTPMGMHSHWTKGKSLSDKEKYAAQATYKKGILPPSPDADYKANPTGD 594
Db 366 -----DSLF-TDPLELEYLYREK-----KKNIDISAK----- 391
Qy 595 SAAATYNVKGKRIPLVRLPYVWHTVEVKNKGLIIPKHQYHNKIFAMPDHYKAPN 654
Db 392 -----VETKEST-----EPNEY--PN 405
Qy 655 GYTLLEDFATIKYVEHDEPHSDNGMGNASEHYLKKHSDENKPKADEPVEETP 714
Db 406 GVT-----YPLSYNDINNALNE--LNSFGDLINFPDYTKESKNIYTDNEKKFIN 454
Qy 715 AEPEVQVETEKVQAQ-----LKEAEVLLAKVTSSLSKANAT---ETLAG 756
Db 455 EIKERIKIEKKIEESDKSYEDRSKSLNDITKEVEKLEINEIYDSKFNNDILTNPEKMG 514
Qy 757 LRNNLTLOIMDNNSMA-----BAEKLIALLK 783
Db 515 KRYSTYKELTHHNTFASYSKSNLEKLTALK 548

RESULT 33
ID USOI YEAST STANDARD: PRT: 1790 AA.
AC P23386;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Intracellular protein transport protein USOI.
GN USOI OR INT1 OR YDL058W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=X2180-1A;
RX MEDLINE=91185402; PubMed=2010462;
RA Nakajima H., Hirata A., Ogawa Y., Yonehara T., Yoda K.,
RA Yamasaki M.;
RT "A cytoskeleton-related gene, usoi, is required for intracellular
RT protein transport in Saccharomyces cerevisiae."
RL J. Cell Biol. 113:245-260 (1991).
RN [2]
RP SEQUENCE OF 782-1790 FROM N.A.
RA Hostetter M.K., Herman D.J., Bendel C.M., McClellan M., Tao N.,
RA Kendrick K.E.;
RN Submitted (Feb-1993) to the EMBL/Genbank/DBJ databases.
RL [3]
RP SEQUENCE OF 1-8 FROM N.A.
RA Bai Y., Symington L.S.;
RL Submitted (May-1996) to the EMBL/Genbank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR PROTEIN TRANSPORT FROM THE ER TO THE GOLGI
CC COMPLEX.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSOCIATED WITH INTRACELLULAR
CC MEMBRANES. PROBABLY PRESENT ON VESICLES OPERATIONAL BETWEEN THE
CC ER AND THE GOLGI COMPLEX.
CC -1- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPEATITIVE, COMPOSED
CC OF AN HEPTAPEPTIDE REPEAT PATTERN CHARACTERISTIC OF ALPHA-HELICAL
CC COILED COILS. MAY FORM FILAMENTOUS STRUCTURES IN THE CELL.
CC -1- SIMILARITY: BELONGS TO THE VDP/USOI/YBL047C FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

```

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X54378; CAA38253.1; -
DR EMBL; L03188; AAB00143.1; -
DR EMBL; U53668; AAB66559.1; -
DR SGD; S0002216; USOI.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR006955; Usol_p115_C.
DR InterPro; IPR006953; Usol_p115_head.
DR Pfam; PF04871; Usol_p115_G1.1_head.
DR Pfam; PF04869; Usol_p115_head; 1.
KM Transport; Protein transport; Golgi stack; Cytoskeleton; Coiled coil.
FT DOMAIN 1 724
FT DOMAIN 725 1790
FT DOMAIN 465 487
FT DOMAIN 991 1790
FT DOMAIN 1172 1786
FT CONFLICT 847 847
FT CONFLICT 924 924
FT CONFLICT 1253 1253
FT CONFLICT 1319 1319
FT CONFLICT 1461 1461
FT CONFLICT 1581 1581
FT CONFLICT 1600 1600
FT CONFLICT 1661 1661
FT CONFLICT 1772 1772
SQ SEQUENCE 1790 AA; 206424 MW; 6CE2B216E9FD4818 CRC64;

Query Match 3.3%; Score 137.5; DB 1; Length 1790;
Best Local Similarity 18.8%; Pred. No. 12;
Matches 153; Conservative 122; Mismatches 254; Indels 283; Gaps 38;

Qy 27 TOKTENTLPBESVREGINAEIVIKITDQGYVTSHGDIHYVNGKPYDAIIEELIMK 86
Db 818 TKDKENQTA-LLEKSTIHKQEDSIKLEKLEI-----ILSQ----- 854
Qy 87 DPNYKLDKEDIIVNEVKGQYIKVYDKGYYVYLKDAHADNVRTKEINRQKQESHQREG 146
Db 855 -----KKKAEDGQINCM-GKDLFALSRMQAVEENCKNLQEKDKSNVNHQKETSLL----- 904
Qy 147 TPRDGAVALARSGRRTTDDGYIFNNSDIIEDGDVYIVHGHYHYPNELS----- 201
Db 905 --KEDIKAKITE-----IKAINENLEEM-----KIQCNNLSKEKEH 938
Qy 202 -ASELAAEAFLSGGNL-----SNSRTYRONSNDTSRTNWVPSVSNPCTTN-TN 250
Db 939 ISKELVEYKRFQSHDLVAKLTREKLSLANNYDMQMAENESLKYAEBESNESSIQLSN 998
Qy 251 TSNSNTNSQASQNDID--SLKQLYKLPISQHVESDGLVDPQAQITSRATGAVPH 308
Db 999 LQNKIDMSQEKENFQIERGSIENKINIEQLKTTISDLQ-----TKKEIISKSDS----- 1048
Qy 309 GDHHPFIPYQMSLEERIRIIPLKRYSNHWVDSRPQSPPTPEPSGPPAPNLK 368
Db 1049 KDEYE-----SQISLKKERLETATTAN----- 1070
Qy 369 IDSNSLVQSOLVRKGVGYFEKIGIRYFAKDLPEE--TVKULESKL-SKQESVHTL 425
Db 1071 -DENVKISETLTK-----RELEAEHLAYKVLKNELEKLETSKAL 1112
Qy 426 TAKKENVAPROEFDYKAYNLLTEAHKALFXNKNRNSDFQALDILRLNDESTNKEKLV 485
Db 1113 KEVEN-----EHLKEKIQLEKATETK-----QQLNSL--RANLESLEKEH-- 1154
Qy 486 DLLAFAPLIPTRLGKPNQSIETDEVRIAQADKYTTSDGYITDDEHDIISDEGDY 545
Db 1155 EDLAAQLK--KYEQIA--NKEROYNE-----ISQNLDELISTQ-----QENESIKKXDEL 1203

```



```

FT DOMAIN 2071 2077 ASP/GLU-RICH (ACIDIC)
FT DOMAIN 2124 2136 ALA/PRO-RICH
SQ SEQUENCE 2175 AA; 233628 MW; 08BF80C461BD0AB CRC64;
Query Match 3.3%; Score 137.5; DB 1; Length 2175;
Best Local Similarity 19.9%; Pred. No. 16;
Matches 125; Conservative 89; Mismatches 244; Indels 169; Gaps 24;
OY 57 GYVSHGHYHYHYNKVPDPAIDISELLMKDPNYLTKDEDIYNEVKGGVIVKDGKYYV 116
DB 254 GSTTSNNAHTSNNSHQ-----DEEIDDEEDEDDEEDEDENSM----- 297
OY 117 LKDAHADNV-----RTKEINROKQESHQREGGPRND--CAVALASQGGYTTDD 167
DB 298 ---OSNADMELDAQOERTERTSPATQOQHOOQDTDELEENKDAEALNSNNHTTDS 354
OY 168 GYFNASDIIEPTGDAYIVPHGDHYHYIYPKELASAEALAEAFSLSGNLSNRTYRQ 227
DB 355 N---NSCSRKNNGG-----NESEQHVASAEDDDCANNNTVTS----- 390
OY 228 NSDNTSRNWPVSQVNPCTTNTS-----NNSNTNSQ-----ASQNDIDSL 271
DB 391 NNNNTSNT---ATSNNTNNNNSSSGNSEKRRKNNNNNGQPAVLAAADKEIKAL 446
OY 272 KOLYK-----LPLSGRHYSDGLVFPDQITRTAGVAVPHGDHYH 313
DB 447 DELQRLAQEOETHLVQIQRLHEHLEVKQCHITRLERARDKQOINALEATRLSAAATN 506
OY 314 FIPYSQSELEERIRIIPLYRYSNHWVPDSRPEOP---SPQPTPEPSPGQAPNLKI 369
DB 507 NNNNSOSSDNKKL-----NTAAERPMADSSNADPESTKAPVPAEDDEE 551
OY 370 DSNLSVLQVLRKVEGVFEKGRVYFADLSET-VKRLSEKLSKQEVSTTL-- 426
DB 552 DEDQML-----VDSEE-----AEDKPEDSHDDEDEDREAVNATTTDS 593
OY 427 ---AKENVARPDQEFYDKAYNLTL-----FAKALFXNKGKNSDFOLDLTLRL 474
DB 554 NELKTKKQOHSPLD-----LNVLSFNSATIAAAAAAACAANDPKFOAL--LIER- 643
OY 475 NDESTNKEKLYVDLLAFAPITHPERLKPNSQIEYTEDEVRIQALADKYTTSDGYFDE 534
DB 644 -TKALAAELKNGASDALSEDAHQOQHOOQHOOQHOOQHOOH--HHHLQQO 699
OY 535 HDIIDEEDGDAYTPRMGSHMWIGKDSLDEKRYAAQATYKKGILPPSPDA----- 585
DB 700 PMSGNSNPASNDHHGH-HLHGHLHPSSAHHLHQTESNSNSTPTAGANNNGSNN 758
OY 586 ---DVKANPTGDSAAIYVRVKGKRI 609
DB 759 SSSNTNANSTAOALASLASTLNGTSL 785

```

```

RT associated protein (MAP1B)-encoding cDNA."
RL Gene 172:307-308(1396).
RP SEQUENCE OF 96-2459 FROM N.A., DOMAIN, AND INDUCTION.
RC STRAIN=Sprague-Dawley; TISSUE=Brain, and Glial tumor;
RX MEDLINE=92347374; PubMed=1639092;
RA Zanner W., Kratz J., Staunton J., Feick P., Wiche G.;
RT "Identification of two distinct microtubule binding domains on
RT recombinant rat MAP 1B."
RL Eur. J. Cell Biol. 57:66-74(1992).
RN [3]
RP SEQUENCE OF 1541-2459 FROM N.A., AND TISSUE SPECIFICITY.
RC TISSUE=Spinal cord;
RX MEDLINE=90059871; PubMed=255150;
RA Rientze A., Gremington G., Hermans-Borgmeyer I., Kirsch J.,
RA Litalner U.Z., Prior P., Gundelfinger E.D., Schmitt B., Betz H.;
RT "Neuraxin, a novel putative structural protein of the rat central
RT nervous system that is immunologically related to microtubule-
RL associated protein 5."
RL EMBO J. 8:2879-2886(1989).
RN [4]
RP DEVELOPMENTAL STAGE, AND PHOSPHORYLATION.
RX MEDLINE=97405699; PubMed=9260743;
RA Ma D., Nothias F., Boyne L.J., Fischer I.;
RT "Differential regulation of microtubule-associated protein 1B (MAP1B)
RL in rat CNS and PNS during development."
RL J. Neurosci. Res. 49:319-332(1997).
CC -1- FUNCTION: The function of brain MAPs is essentially unknown.
CC Phosphorylated MAP1B may play a role in the cytoskeletal changes
CC that accompany neurite extension. Possibly MAP1B binds to at least
CC two tubulin subunits in the polymer, and this bridging of subunits
CC might be involved in nucleating microtubule polymerization and in
CC stabilizing microtubules.
CC -1- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
CC with MAP1A and MAP1B proteins.
CC -1- TISSUE SPECIFICITY: Nervous system (spinal cord, brain stem,
CC cerebellum and cerebrum). Not expressed in liver, spleen, kidney,
CC heart or muscle.
CC -1- DEVELOPMENTAL STAGE: In cerebral cortex, spinal cord and sciatic
CC nerve levels are high early in development but decrease during
CC postnatal development and are low in adults. In dorsal root
CC ganglia levels remain high throughout development.
CC -1- INDUCTION: By nerve growth factor.
CC -1- DOMAIN: Has a highly basic region with many copies of the sequence
CC KKEE and KKEL/V, repeated but not at fixed intervals, which is
CC responsible for the binding of MAP1B to microtubules.
CC -1- PTM: LC1 is coexpressed with MAP1B. It is a polypeptide generated
CC from MAP1B by proteolytic processing. It is free to associate with
CC both MAP1A and MAP1B. It interacts with the amino-terminal region
CC of MAP1B (by similarity).
CC -1- PTM: Phosphorylated.
CC -1- SIMILARITY: TO MAP1A.
CC -1- CAUTION: A C-terminal fragment of this protein (residues 1597 to
CC 2459) was originally described as neuraxin in Ref.3.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U52950; AAB17068.1; -
CC EMBL; X60370; CAC16162.1; -
CC EMBL; X16623; CAA34620.1; ALT_SEQ.
CC PIR; A56577; A56577.
CC InterPro: IPR000102; MAP1B_neuraxin.
CC Pfam; PF00414; MAP1B_neuraxin; 10.
CC PROSITE; PS00230; MAP1B_NEURAXIN; 8.
CC Microcubules; Repeat; Phosphorylation.
FT CHAIN ? 2459
FT REPEAT 1869 1885 MAP1B 1.

```


FT DOMAIN 23 308 FERM.
 FT ACT SITE 923 1176 PROTEIN-TYROSINE PHOSPHATASE.
 FT DOMAIN 1110 1176 BY SIMILARITY.
 FT DOMAIN 340 343 POLY-PRO.
 FT DOMAIN 565 572 POLY-PRO.
 SO SEQUENCE 1176 AA; 133490 MW; 529FBE22F1335B75 CRC64;

Query Match 3.3%; Score 137; DB 1; Length 1176;
 Best Local Similarity 19.1%; Pred. No. 7.2; Indels 222; Gaps 32;
 Matches 137; Conservative 90; Mismatches 267;

QY 29 KTEMLTPEVSKREGINAEQIVIKITDQYVTS-----HGHHYHNGKVPY----- 75
 DB 202 EAEILLYMGEVERMDQVGEESYPK--DSQSDSISIGACLDGIVKKNQRPVVRMHDIA 260
 QY 76 ----DAIIEELMKDPNYKIKDEDYNEVGKVIYKDKYIYLLKDAHADVNRKE 130
 DB 261 NMSHNKSPFALELANKEETICFOTEDMET-----AKYVNRLCVARKHFYRLNQC 309
 QY 131 EINRQKQSHSCHRREGTGRNDGAVALARSO-----GRYTTDDGIIFNASDII 177
 DB 310 SLQTPAATLNSVYRDSSSR---MSLPKQPYAMPPPQLHYNGHYT-----BPF 355
 QY 178 EDTGDYIVPHGDHYIIPKXELASAEELAAEAFISGRGNLSNRTY--RQNSDNTSRT 235
 DB 356 ASSQDNIIVPNKNGFYCHSQISLDRQID---LSGR--IRNGSVGAHSTNSLNTIQP 408
 QY 236 MNVPS--VSNPCTTNTSNNSTNSQASQ-----SNDIDLLKQLYK-LPLSQRH 283
 DB 409 YLQPSMSSNPISITSDVWRPDSLPSHRSHALIPSYRPTPYETVVMKLNKGWVAHRH 468
 QY 284 VES-----DGLVDPAQITSTRAGAVAPHGDHYF-----IPYSQ 319
 DB 469 SHSLANLIGSSYASRPDALVYSQPEI--RHPHLTSPQSAHYFPNLNYSFHSQSPYR 526
 QY 320 MSELEERLARIIPLYRSNNHWVDSRPE-----QSPPOPT-----PEPSGPQ 362
 DB 527 PAERPPVYGAV-----SVPELTNVQLAQDYAPANIMRTQVYRPPRYPR 573
 QY 363 PA---PNLK-----IDNSSLVSQVLRKYGEGYEEEGKISRYVAKDLPSTTVKNLSK 414
 DB 574 PANSTPDLISRLHYISSNPDLITRRVHNSVQ--TFQEDSLPYAHSLEGESEPLTAARAH 631
 QY 415 LSKQSV-----SHTLTAKKNVAPR-----DOEFYDKAVNLTLE--- 449
 DB 632 LQKRISITAGLTGPEGRLKERTVSASADVAFRTSAGSQSSVSFQKMKQESTBQOE 691
 QY 450 ----AHKALFXNKGNSDFQALDKLERLNDESTNKEKLVDDL-----L 489
 DB 692 GGRVSHKKSLSDPATMLIDSEDEDELE---EDSSREQAISAVSEPRLLTAAFSOSLNYPC 747
 QY 490 AFLAPITPERLKGKNSQIETEDEVRIALQADKYTSDGYIFDEHIIISDEGAYVTPH 549
 DB 748 ASAPFITGPHLIFEKPPHY--TEPEKRAKDIDSPVH-----LVETH---RPRDGLLTP 796
 QY 550 MGHSHWIKDLSDEKXVAAQAYTEKEGILPPSPADYKANTPTGSAAIYNRVKG 605
 DB 797 -----SMSESBDLTTSGRYRARD-----SVKKRVSDLLSGKSAVEG 834

RESULT 37

Y153 YEAST

ID Y153 YEAST STANDARD; PRT; 679 AA.

AC P40563;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Hypothetical 74.8 kDa protein in BET1-PAN1 intergenic region.
 GN Y1R003W OR Y1B3W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OC NCBI_TaxID=4932;

RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=95282515; PubMed=7762303;
 RA Voss H., Tamames J., Teodoru C., Valencia A., Senses C., Wiemann S.,
 RA Schwager C., Zimmermann J., Sander C., Ansgore W.;
 RT "Nucleotide sequence and analysis of the centromeric region of yeast
 RT chromosome IX.";
 RL Yeast 11:61-78(1995).

RN (2)

RP SEQUENCE FROM N.A.

RC STRAIN=S288C / AB972;

RX PubMed=9169870;

RA Chircher C.M., Bowman S., Badcock K., Bankier A., Brown D.,

RA Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,

RA Harris D.B., Horsnell T., Hunt S., Jagels K., Jones M., Lye G.,

RA Mould S., Odeli C., Pearson D., Rajandream M.A., Rice P., Rowley N.,

RA Skelton J., Smith V., Walsh S., Whitehead S., Bartell B.G.;

RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX.";

RL Nature 367:84-87(1997).

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; X79743; -; NOT ANNOTATED_CDS.

DR EMBL; Z38062; CAAB6205.1; -;

DR PIR; S48437; S48437.

DR SGD; S0001442; Y1R003W.

KW Hypothetical protein.

SQ SEQUENCE 679 AA; 74763 MW; 9DF79500375339E7 CRC64;

Query Match

3.3%; Score 136.5; DB 1; Length 679;

Best Local Similarity 19.3%; Pred. No. 3.6;

Matches 144; Conservative 97; Mismatches 260; Indels 247; Gaps 33;

QY 121 AHADNVTKEERINRQKQSHSCHRREGTGRNDGAVALARSOGRYTTDD--GYIFNASDII 178
 DB 42 AGTNVPTRRRLILAKTMTS--FESGMDQESLPKVPLOPRARSTTEELNNMNTSKYLE 100
 QY 179 DTGAYIVPHGDHYIIPKXELASAEELAAEAFISGRGNLSNRTYRQNSDNTSRTNW 238
 DB 101 EI-ESLISKNIHIVSRKSPSYVEEGKVAIHONG-----CRSADNKTSTNPS 149
 QY 239 PSVEN-----PGTNTNTSNV--SNTNSQASQNSDIDSLKQLYKLPISQGRHVESDGLVF 291
 DB 150 PLEKNEHEGAGNESAISPSNLVKNNSNEVTEHDSDEDLTEK-----QKYHALDNBAG 203
 QY 292 DPAQITSRAGVAVPHGD-----HYHPIYSQNSELEERLARI-----PLRYRSN 338
 DB 204 DRSHFEKRLIPGDKVQDVDSKDVEEGSLNALPSPGITESDDKAKFTKPPESLLELOK 263
 QY 339 HNVVDSRPEQSPPTPEPS-----PGQPPANLKDINSLSVLSQ 378
 DB 264 H--OEQOEKIFQNPTEESTSTLSNEKQEGKNNEVNSQPGPDTETVIAATSNVPSQ 321
 QY 379 LVKRVGGYVEEGGIRGVYFAKDLP-----SETYKNLSKLSKQESVSHLTAKKNVAP 434
 DB 322 IAS-----EEE-----NDVPVIPSRRPKDPEAHVQKE-----LP 352
 QY 435 RDOEFYDKAVNLTLEAHKALFXNKGNSDFQALDKLERLNDE-----STNKEKLVDDL 488
 DB 353 NTQF-----KVSSECDSTLSTEEESKIPKI 379
 QY 489 LAFAPITPERLKGKNSQIETEDEVRIALQADKY-----TSDSYIFDEH 535
 DB 380 -----PSERPRAAPVPKPPSSRIAFQEMLQKQOQOQDLHNNGNSATTASADIAKKH 434
 QY 536 ---DIISDEGDAYVTPHMGSHWIGKDSLSDKEKVAQAAYTEKEGILPP-----SPDA 585

```

Db 435 TDSSTSDTTKA-----DFTSKLNGFLALPGMVPQGLPSPLEKLSPT 480
OY 586 DYKANPTGDSAAIYNRVKGEKRIPL-----VLPVVEHTVEVKONLIIHRK 634
Db 481 ESKLGPDOQO-----KTGPGTRRGCGPRGRLPSKVASVEKIE----- 523
OY 635 DHYHNIKFAMFDDHTYKAP--NGYTLLEDLPATIKYVEHPDERPHSNDGMNASEHVLGK 692
Db 524 -----DNTNKIKIIFNNMNVSSSFSEKXVLI---DITP-----GGAERALDE 563
OY 693 KDH-----BDPNK-----NFKADEPVEETPAEPVPOVEYTERKVAOLKEAEVLLA 739
Db 564 KSKSIPEEQREQSPNMEALCPFEIDEX--EKLPAHAESDPL-----SQLPOTNAVGN 615
OY 740 KVTDSLSKANATETLAGLNNLTIQIMD 767
Db 616 KRAISEBISPEALANRDQNDTTEIOE 643

RESULT 38
CNA_STAAT
ID CNA_STAAT STANDARD: PRT, 1183 AA.
AC 053654;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Collagen adhesin precursor.
GN CNA.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FDA 574;
RX MEDLINE=92165839; PubMed=1311320;
RA Patti J.M., Jonsson H., Guse B., Witalski L.M., Wlberg K.,
RT Lindberg M., Hoeek M.;
RA "Molecular characterization and expression of a gene encoding a
RT Staphylococcus aureus collagen adhesin.";
RL J. Biol. Chem. 267:4766-4772(1992).
RN [2]
RP ERRATUM.
RA Patti J.M., Jonsson H., Guse B., Witalski L.M., Wlberg K.,
RL Lindberg M., Hoeek M.;
RN J. Biol. Chem. 269:11672-11672(1994).
RN [3]
RP COLLAGEN-BINDING DOMAIN.
RC STRAIN=FDA 574;
RX MEDLINE=94032261; PubMed=8218209;
RA Patti J.M., Boles J.O., Hoeek M.;
RT "Identification and biochemical characterization of the ligand
RT binding domain of the collagen adhesin from Staphylococcus aureus.";
RL Biochemistry 32:11428-11435(1993).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 169-318.
RX MEDLINE=97475225; PubMed=9334749;
RA Smerly J., Patti J.M., Carson M., House-Pompeo K., Teale M.,
RA Moore D., Jin L., Schneider A., Delucas L.J., Hoeek M.,
RA Naryana S.V.L.;
RT "Structure of the collagen-binding domain from a Staphylococcus
RT aureus adhesin.";
RL Nat. Struct. Biol. 4:833-838(1997).
CC -1- FUNCTION: MEDIATES ATTACHMENT OF STAPHYOCOCCAL CELLS TO
CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
CC an amide bond (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M81736; AAA20874.1; -
DR PDB: 1AMX; 24-JUN-98.
DR PDB: 1D2O; 27-SEP-00.
DR PDB: 1D2P; 27-SEP-00.
DR InterPro: IPR001899; Gram_pos_anchor.
DR InterPro: IPR006192; LPXTG.
DR TIGRfams: TIGR01167; LPXTG_anchor.
DR PROSITE: PS00847; GRAM_POS_ANCHORING; FALSE_NEG.
KW Cell wall; Peptidoglycan-anchor; Repeat; Signal; 3D-structure.
FT SIGNAL 1
FT CHAIN 30 1154
FT PROPEP 1155 1183
FT DOMAIN 151 318
FT DOMAIN 533 1093
FT DOMAIN 1093 1157
FT REPEAT 533 719
FT REPEAT 720 906
FT REPEAT 907 1093
FT SITE 1151 1155
FT MOD_RES 1154 1154
FT STRAND 174 179
FT TURN 182 183
FT TURN 185 186
FT STRAND 187 194
FT TURN 196 197
FT STRAND 201 201
FT STRAND 205 211
FT STRAND 215 228
FT TURN 229 230
FT STRAND 232 234
FT TURN 239 246
FT TURN 248 249
FT STRAND 251 255
FT TURN 256 259
FT STRAND 260 265
FT HELIX 267 270
FT TURN 271 272
FT STRAND 273 283
FT TURN 286 287
FT STRAND 290 299
FT STRAND 301 301
FT TURN 302 303
FT STRAND 307 311
FT STRAND 314 317
SQ SEQUENCE 1183 AA; 133066 MW; 6A1CC072E575D76 CRC64;

Query Match 3.3%; Score 136.5; DB 1; Length 1183;
Best Local Similarity 19.2%; Pred. No. 7.7; Indels 357; Gaps 54;
Matches 189; Conservative 120; Mismatches 317;

OY 15 NNRVSYIGKQKOTQNTULPDEVKREGINAEQIVIKTDQGYTSHGDHYHYNGKVP 74
Db 135 NNEKRYV-SKOTTK-----DQIGGQGLDSTININVT-----GTHSNYISGQ-- 237
OY 75 YDAIISSELLMKDPNYKDEDIYNEVK-----GGY-----VIKVDGKYYVYL 117
Db 238 -SAIDFEKAF--PSKITVDNTKNTIDVTIPQGYGVNSFSINYYKTKITNEQKEFVNN 294
OY 118 KMAHADNVRTKEEINNRQKQEHSCHEGTRPNDGAV-----ALARSQGR 162
Db 235 SQAWYQEH--GKEEVNGKSFNHTVINANAGIEGVGKELVKLQDKDTKAPIANVKPK 352
OY 163 YTTDDGYGFNAS-----DIIED-TGDAYI--VPHGQHYHYIPKNELSASELAAEFLSGR 215
Db 353 LSKKGSGVKNQKQKEIETITDANGIANKALPSCG---YILK-----EIEAPRYTPPK 403
OY 216 GNLSRTRYRONSNTSRTNKPVSNSPGTTNTNNSNTNSQASQSGNDIDSLLKOLY 275
Db 404 ---DREYFTWMDTQNGYF-----TTIENAKAIETKVDVSAQK----- 439

```

```

Oy 276 KLPLSQRHVESDGLVFDPDAQITSRTARGAVAPVGHGHHFIPY--SQMSELEERIAIPL 333
Db 440 -----VMEGTQVKPPTIYFKLYKQDNQNTTPVDKAEIKLEGGTTKV-- 483
Oy 334 RYRSNHWVPDGRPEPSPQPTPEPSPGPAPNLIKIDNSSLSVLQVKNV-----E 385
Db 484 -----W--SNLPEN-----DKNGKAIKLYEVANAGEDTTPPE 514
Oy 386 GYVFEKGISRYVFAKDLPSFTVKNLLESL-----SKOESYHTLTAKENAVAPRD 436
Db 515 GYTKKENL--VYMTTEKPIETTSISGEKWDKDNQDGKREKXSVNLLANGKVKTL 572
Oy 437 -----QEF-----YDKAVNL--LTFAH-----KALFYXK-- 458
Db 573 VTSETNMYEYFADLPKYDEGKIEYTVEDHVKQYTTDINGTTNTKTPGETSATVTKN 632
Oy 459 ---GRNSPQALDKLERL--NDESTNKEVLDLAFAPLTPHERLGRKPSQIETED 513
Db 633 WDDNNNODGKRPTELKVELYODGKATGTALINENNNHTWTGDEKAK--GQGVKYTVE 691
Oy 514 EVRIAQLADKYTSDGY--PDEHD-----ISDEGDAVTPHMGSHWIGKDS----- 560
Db 692 EL-----TKVAGYTHVDNNDMGNLIVNTKTPETTSISGEKWDKDNQDGKRP 742
Oy 561 -----LSDEKQVAAQAYTKEKILPSPDADVKANPTGDSAAAIYNNKGEKRIPLVRL 614
Db 743 KVSVNLADGEKVKTLDTVSE-----TNMKYEFKLPKYD-----EGKIEYVTE 788
Oy 615 PYMVEHTYEVKKNGLI-----IPHKHYHN-----TPAMFDD----- 647
Db 789 DHVADYTTDI--NGTITNTKYTPGETSATVTKNWDNNNODGKRPTEIYVELYQDGKATGK 847
Oy 648 -----HTY-----KAPN-----GYTLEDLPATIKYVVEPDERHSNDGWN----- 684
Db 848 TALINENNNHTWTGDEKAKGQGVKYTVEEL--TKVAGYTHVD-----NNDMGNLIVT 901
Oy 685 -----ASEHVLGKKDSED--PNK--NFKADEPVEETPAEPE-----VP 720
Db 902 NKYTEPETSISGEKWDKDNQDGKRPKVSVNLLANGKVKTLVSETNMYEYFADLP 961
Oy 721 QV-EFEKVEAQLKEVLLAKYTDSSLKANAT-----ETLAGLNNLTLOIMDNN 770
Db 962 KYDEBKKEIYVTEDEH-----VKDYTTDINGTTITNTKYTPGETSATVTKND--DNNN 1012
Oy 771 IMAEAKULLALLKGSNPSSVSKE 793
Db 1013 -----QDGKRPTEIKVE 1024

```

```

RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -I- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (potential).
CC -I- PM: MEROZOITE SURFACE ANTIGEN CONTRAIN THE SEQUENCE OF 83 KDa, 42
CC KDa AND 19 KDa ANTIGENS WHICH ARE THE MAJOR SURFACE ANTIGENS OF
CC MEROZOITES. THE MATURATION TAKE PLACE DURING SCHIZONT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03371; CA27070.1; -.
DR InterPro; IPR006209; EGF_like.
DR Pfam; PF00008; EGF; 1.
KW Malaria; Merozoite; Polyprotein; Repeat; signal; Glycoprotein;
KW Transmembrane; GPI-anchor.
FT SIGNAL 1..19 POTENTIAL.
FT CHAIN 20..1630 MEROZOITE SURFACE PROTEIN 1.
FT DOMAIN 67..84 TRIPEPTIDE SG(TP) REPEAT.
FT TRANSMEM 1614..1630 MEMBRANE ANCHOR.
FT CARBOHYD 97..97 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 259..259 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 755..755 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 759..759 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 774..774 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 835..835 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 911..911 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 955..955 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1049..1049 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1156..1156 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1165..1165 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1436..1436 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1517..1517 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1630 AA; 187289 MW; ADBDEC3CE0A46322 CRC64;

Query March 3, 34; Score 136; DB 1; Length 1630;
Best local similarity 19.84; Pred. No. 13;
Matches 132; Conservative 90; Mismatches 229; Indels 216; Gaps 29;

Oy 186 VPHGDHYHYIPKNN-----LSASELAARAFLSGRGLNSRTYRRONSNTSRITNV- 238
Db 20 VTHSYQELVKKLEALDAVLTVGSLPHKEMILNEBEITTKGASQSGTSGTSGTSPS 79
Oy 239 -PSVSNPGTNTNTNNSNTNSQASQNDI--DSLKKQI-----YKLPISQRHVESD 287
Db 80 GPSGTSF--SSRSNTLPNSNTSSGASPPADASDSAKSVADLKHVRNVLTLTIKELKYPQ- 137
Oy 288 GLVFDPDAQITSRTARGAVAPVGHGHHFIPYSQMSLEERIAIIP--LRYRSHNVVPDSRP 346
Db 138 --LFD--LNNHMLTLCDNIHGFRYLIDGYEINELLYKLFYDPLAKAKNDVACANDYC 192
Oy 347 EQPQPPTPEPSPGPAPNLIKIDNSS--SLVSQV--RK-----VG--EGVFEK 392
Db 193 QIPF-----NKKIRANELDVAIKLVPGRKRLDNIKONVGMEYIKKNK 237
Oy 393 GISRYVFAKDLPSFTVKNL-----ESK--LSKQESVHTLTAKENAVAPRQDEFYDK-- 442
Db 238 -----KTIENINELIESKKTIDKNKATKEEKKLYQAOYLSIYNKOLE 284
Oy 443 -AYNULTFAHKAALFXNKGNSDPALDKLERLNDESTNKEVLDLAFAPLTPHERL 501
Db 285 EAHNLISVLEKRIDTLTKNENIKELDKINIKNPPANSNT-----PNTL 331
Oy 502 GKPSQIETEDYEVRIAQLADKYTSDGYIPDEHDIISDEGDAVTPHMGSHWIGKSL 561
Db 332 LDKKKKIBEHKEKEIKETAKTIKFN-----DSL 359
Oy 562 SDEKQVAAQAYTKEKILPSPDADVKANPTGDSAAAIYNNKGEKRIPLVRLPYMVEHT 621

```

Db 360 F-TDPLEEYVUREK-----NKNIDISAK----- 382
 Qy 622 VEVKGNLIIIPKDHYNHIFAMFDDHTYKAPNGYTLLEDFATIKYVEHPDERPHSNDG 681
 Db 383 VETKST-----ERNEY--PVGVT-----YPLSYNDINNALNE 413
 Qy 682 WGNASEHVLGKKDHSDDPNKFNKADEEPEETPAEDEVPOVETEKVAQ----- 730
 Db 414 -LNSFGDLINPFDYKPEPSKNIYTDNERKKFINEIEKIKIEKKKIESDKSKYEDRSKSL 472
 Qy 731 ---LKAEEVLAKVNTSSLSKANAT---ETLAGLRNNLTIQIMDNNSIMA-----EAE 776
 Db 473 NDIKYEKELNEMIBYDSKFNNNIDLTNFEKMGKRYSYKVEKLTHTNTFASYENSKHLE 532
 Qy 777 KLLALAK 783
 Db 533 KLTALK 539
 RESULT 40
 BUD3 YEAST STANDARD; PRT; 1636 AA.
 ID BUD3 YEAST STANDARD; PRT; 1636 AA.
 AC P25558, P25557, P87007;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Bud site selection protein BUD3.
 GN BUD3 OR YCL014W OR YCL14W/YCL13W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95247824; PubMed=7730410;
 RT Chant J, Mische M, Mitchell E, Herskowitz I, Pringle J.R.;
 RL "Role of Bud3p in producing the axial budding pattern of yeast.";
 RL J. Cell Biol. 129:767-778 (1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX Oliver S.G., Anwar R., Brown A., Gent M.E., Indge K.J., James C.M.,
 RL Staveva L.;
 RL Submitted (MAR-1992) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP REVISIONS.
 RA Gromadka R.;
 RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP REVISIONS.
 RA Valles G., Voelckers G.;
 RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: CO-ASSEMBLES WITH BUD4 AT BUD SITES. BUD4 AND BUD3 MAY
 CC COOPERATE TO RECOGNIZE A SPATIAL LANDMARK (THE NECK FILAMENTS)
 CC DURING MITOSIS AND THEY SUBSEQUENTLY BECOME A LANDMARK FOR
 CC ESTABLISHING THE AXIAL BUDDING PATTERN IN G1.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; U1580; AAA6315.1; -;
 DR EMBL; X59720; CAA42346.2; -;
 DR PIR; S74285; S74285.
 DR SGD; S0000520; BUD3.
 DR InterPro; IPR000219; RhogEF.
 DR SMART; SM00325; RhogEF; 1.
 KW Cell cycle.
 SQ SEQUENCE 1636 AA; 184717 MW; 9E4E6BA5C3A3F69 CRC64;

Query Match 3.3%; Score 136; DB 1; Length 1636;
 Best Local Similarity 19.6%; Pred. No. 13;
 Matches 171; Conservative 109; Mismatches 321; Indels 270; Gaps 42;
 Qy 22 DGRQATQKTENITPPEVSKREGINAEQIYIKITDQGYTSHGDHYHNGKYPDAIISE 81
 Db 798 DMSNVTYKETEKLIPVKPTKSSKKPREI-----QKTKTNASKAE 837
 Qy 82 ELMKDPYKIMDEDIIVAEVKGYIYVKGK-----YVYLKDAHADNVRTTEELNRQ 137
 Db 838 HIEKKPN--KGKGFPGVLKVPFSKSKSPSPVQRPVKITISQHPSPVKPKPMTSEKK 894
 Qy 138 EHSQREGEPT--RNDGVALARSGRYTTDDGIIFNASDIIEPTGDAIYIPHDHYH 194
 Db 895 SSPKRAVSSPKIKKKTSFSFKESQTKAKSLRAVEFSDDLIGRPD--VGNAGH--- 948
 Qy 195 IPRNELSASELAALAEPLSGRNLNSRTYRRQNSDNTSRINWVPSVSPGTTNTNTSN 254
 Db 949 -POENTRISSVVDRTKYVS--YNPQPYT-----ENTSNEKVEPKADQSTKODNISNF 999
 Qy 255 SNTNSQAS-----QSNDDISLKLQYKPLSQRHVESDGLVFPDPAQITRTAGVAV 306
 Db 1000 ADVENASVPEPKLDAETD-----QITGKATNSSVHGKNEPLDAEVT--TANRVST 1051
 Qy 307 PH-GD-----HYFLIPYSQMSL-----EERI-----ARIPLRSHHWVPSDR 345
 Db 1052 TSAGDQRIDTQSEFLRAADVENVLSDDDEHRQNSKRVFNDLRCDFIPKHYRK----- 1104
 Qy 346 PEQSPQPTPEPSFGPQAPANIKIDSNSGLVSQLVRKVGEGYVEEKGISRYVFAKDLS 405
 Db 1105 -----QENINSSNLPF-----EGKVPQEGVSN----- 1128
 Qy 406 ETVKNLSEKLSQESVSHITLTKKENVAPRODEFDKANULTTEAHKALFXKKGKNSDQ 465
 Db 1129 ---ENTNISLKTNEAS--TLT-----QKLSQASKVLTESSNELKOTNNECKDAR 1174
 Qy 466 ALDKLERLNDSTYKNEKLVLDLFLAPITHP-----RLGKPNQIEYTEDEVRIA 518
 Db 1175 DI-KIGDDYSKETAKE--ITPKNFVEGITRKEKIFPIPIPLAPASKINQR----- 1225
 Qy 519 QIADRYTTSQGYI--FDEHDIISDEGDVAVTPHMGSHW--IGKDSIDKEKVAQAAYT 573
 Db 1226 -----SPSYIEIFQGWKRVVLDKHA-----HYWKRILASQVSLSEGLKVTE--- 1267
 Qy 574 KEKGILPSPDADVYAN-----PT-----GSA-----A 597
 Db 1268 EDAAIINKSQD--DAKAERMTQISEVIEYEMOQPIPTLPKALHDSGIEKSDKFEIIE 1326
 Qy 598 AIYNRVKGEK-----RIPLVRLP--YVWEHVEYKGNLIIIPKDHYN--- 639
 Db 1327 ELKEELKSKTGNEDVGNPNPNSIPKIEKPPAFKYIRTSPVRIIGRTFEDTRKYENGSP 1386
 Qy 640 --IKFAMPDDHTYKAPNGYTLLEDFATIKYVEHPDERPHSNDGNGNASEHVLGKKDHS 697
 Db 1387 SDISFT-YDTHNNDPDKRLMELKFPSS--QGEIFDDRYT-----PAE 1426
 Qy 698 DPNKFPKADEEVEETPAEDEVPOVET--EKVEAQLKAEV-----LLAKYDSSLSLA 748
 Db 1427 EPTAEF-----PVEELPNTPRISINVTTSNNKSTDKLSGNDIKETELDLLEFSSFNI 1481
 Qy 749 NATETIAGLRN--NLTIQIMDNNSIMAEKEL 778
 Db 1482 AFGNTSMSTDMNKISSDLSNKTIVLGNQKV 1512
 RESULT 41
 SLPN_BACBR STANDARD; PRT; 1053 AA.
 ID SLPN_BACBR STANDARD; PRT; 1053 AA.
 AC P06546;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Middle cell wall protein precursor (MWP).

OS Bacillus brevis (Brevibacillus brevis).
 OC Bacteria: Firmicutes; Bacillales; Paenibacillaceae; Brevibacillus.
 OX NCBI_TaxId=1393;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=47;
 RX MEDLINE=88115203; PubMed=2828336;
 RA Tsuboi A., Uchihashi R., Adachi T., Sasaki T., Hayakawa S., Yamagata H.,
 RA Tsukagoshi N., Uda S.;
 RT "Characterization of the genes for the hexagonally arranged surface
 RT layer proteins in protein-producing Bacillus brevis 47: complete
 RT nucleotide sequence of the middle wall protein gene.";
 RL J. Bacteriol. 170:935-945(1988).
 RN [2]
 RP SEQUENCE OF 1-199 FROM N.A.
 RC STRAIN=47;
 RX MEDLINE=87137282; PubMed=3029027;
 RA Yamagata H., Adachi T., Tsuboi A., Takao M., Sasaki T.,
 RA Tsukagoshi N., Uda S.;
 RT "Cloning and characterization of the 5' region of the cell wall
 RT protein gene operon in Bacillus brevis 47.";
 RL J. Bacteriol. 169:1239-1245(1987).
 RN [3]
 RP SEQUENCE OF 676-1053 FROM N.A.
 RC STRAIN=47;
 RX MEDLINE=87008404; PubMed=2428810;
 RA Tsuboi A., Uchihashi R., Uda S., Takahashi Y., Hashiba H., Sasaki T.,
 RA Yamagata H., Tsukagoshi N., Uda S.;
 RT "Characterization of the genes coding for two major cell wall
 RT proteins from protein-producing Bacillus brevis 47: complete
 RT nucleotide sequence of the outer wall protein gene.";
 RL J. Bacteriol. 168:365-373(1986).
 RN [4]
 RP SEQUENCE OF 1-50 FROM N.A.
 RC STRAIN=47;
 RX MEDLINE=90078123; PubMed=2512285;
 RA Tsuboi A., Uchihashi R., Engelhardt H., Hattori H., Shimizu S.,
 RA Tsukagoshi N., Uda S.;
 RT "In vitro reconstruction of a hexagonal array with a surface layer
 RT protein synthesized by Bacillus subtilis harboring the surface layer
 RT protein gene from Bacillus brevis 47.";
 RL J. Bacteriol. 171:6747-6752(1989).
 CC -1- FUNCTION: THE MIDDLE WALL PROTEIN BINDS TO PEPTIDOLYCAN AND TO
 CC THE OUTER CELL WALL PROTEIN.
 CC -1- SUBUNIT: THE MIDDLE CELL WALL LAYER IS COMPOSED OF SUBUNITS OF
 CC THE MIDDLE CELL WALL PROTEIN. THESE PROTEINS FORM A HEXAGONAL
 CC ARRAY WITH A LATTICE CONSTANT OF 14.5 NM IN THE MIDDLE CELL
 CC WALL LAYERS.
 CC -1- SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A
 CC HEXAGONAL S-LAYER.
 CC -1- SIMILARITY: Contains 3 S-layer homology (SLH) domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib.ch).
 CC -----
 CC EMBL: M15364; AAA87321.1; -;
 CC EMBL: M14238; AAA22372.1; -;
 CC EMBL: M19115; AAA22760.1; -;
 CC EMBL: M31828; AAA22619.1; -;
 CC PIR: A28555; A28555;
 CC InterPro: IPR001119; SLH.
 CC Pfam: PF00395; SLH; 2.
 CC PROSITE: PS01072; SLH DOMAIN; 2.
 CC Cell wall; S-layer; Signal; Repeat.
 KW SIGNAL
 FT CHAIN 1 23 MIDDLE CELL WALL PROTEIN.
 FT DOMAIN 24 1053 SLH 1.
 FT DOMAIN 93 143 SLH 2.
 FT DOMAIN

FT DOMAIN 144 203 SLH 3.
 SQ SEQUENCE 1053 AA; 117146 MW; DB421318BD9D5E4F CRC64;
 Query Match 3.3%; Score 135.5; DB 1; Length 1053;
 Best Local Similarity 18.0%; Pred. No. 7.4;
 Matches 162; Conservative 132; Mismatches 292; Indels 315; Gaps 44;
 QY INASQIVYKIDTDOG-----VYTSKGDHYHYNGKVPYDAIISFELMKDPYKL----- 92
 DB 252 IKANEVLINGDAGIGNTTYVADGINANDPDG-----HYQVWKDKDEDIYVMKEG 304
 QY -KDDIVNEVKGVIYKVDGKYVYLLDAAHADNVRTKEEINRQKQESHQSEGTPEPN- 150
 DB 305 STDQEVIMDRVGEFTLK--GKTFEDPKDLSNLDLKLLEDASEKSYRFMKNKTKVTYNF 362
 QY 151 -----DGAVALARSQGRYTTDDGYIFNADIIEDPCD--AYIYPHGDHYIYIPKNEISA 202
 DB 363 TRENDPVDGLKEIKD-----NADGFTFGAKVLDNNELAYI-----HVIDQSNK 411
 QY 203 SELAAEAFLSGRGNLSRTRYRQNSDNTSRTNWVPSVSNPCTTNTYNSNNSNTNSQAS 262
 DB 412 ED-----EGVKGSEVIS-----KIDTDKKKITTNRND 439
 QY 263 QSNIDISLLKQLYKLPLSQRVESDGLVF--DPAQ--ITSRTARQVAVPHGDHYHIFP 316
 DB 440 KFNLDLG-----KEEGKDFLVFLNGKPAKFSDLKEGVVSYVYADGDEDKLLV 487
 QY 317 YSQNSELEERARIILPLRYSNHNVPPSRPQSPQPTPEPSPQPPAPNLIKIDNSSLV 376
 DB 488 FATVTVGEGKDKVVS--RNN--DYR-----LTI 513
 QY 377 SOLVRKVGEGVYFEKSG-----ISRYPFAKDLPEETVK-----NLESKLS-- 416
 DB 514 GDKTRYVEGATFSDDGKNDYQDIDKDHMDLVSDDETFKLYLDASGRVNHITKDAID 573
 QY 417 --KQESV--SHTLTAKENAVPRDQEFYDAVNLITFAKAL-----F 455
 DB 574 DRKQALVTRSATNTSKDT-----WD--FVLTKQKEITVLSLEAKNIYDFDGKNF 623
 QY 456 XNKRNSDPQADLRLERINDESTNKEKLVLDL--LAFAPLTHPELGRPNQIE 509
 DB 624 SRDKNDP--DLEDILVPSKDKOTLLLEVTLADAGKQKVEFLPVKVEQSGKAMDOLA 681
 QY 510 YTEDEV-----RIAQLADKYTTSDGYIFDEHDI-----ISDE 541
 DB 682 DEDDMVGDIYVTKTAVFNTGKLESSKKEKAKTAKFQVADENDSLVITYVNDK 741
 QY 542 -----GDAYVTTPHNGSHWIGKSL-----SDKEVVAQAQYTKEGILPPS 582
 DB 742 DEVEAIFVVEGDGLTGDPAHYQVVIDFGKGGKDTIRVWEKDGDKVEKEY--KLDDQDDL 800
 QY 583 PDADVKNP-----TGDS-----AAATYNRVKGGRITVLVLPVVEHTV 622
 DB 801 KDEDIRRDFIAFTVDSDEVDVVEVNVKNKAGMLAEVTDKGMKDANIIDKKVGLV 860
 QY 623 -EVKNGNLIIPHKDHYHNKFA-----WFD-----DHTYKAPNG----- 655
 DB 861 SDVKQDT--ITYKADNKKKASIKSATVYDLYDDPGEAGVNGEYVVMIDSGDISGT 918
 QY 656 ---YTL-----EDLPATIKYVEHDERPHSNDGNGNSEHVLKKDHS 698
 DB 919 KYDVVLIVSDAKTVRKDKLEDDDAEAPLK--QESEKPDPTKWDALPSKVEGKFT--SAG 974
 QY 699 PNKAFKADPEVEETPAPEVPOVE--TEKVAQL--KEAVLLA-----KYTSSSL 746
 DB 975 PVKLYRATVELNSKVKA--DVDAIEFYNGKRVPSLNFQDGVITGYNTEDKVTSSKI 1033
 QY 747 K 747
 DB 1034 K 1034
 RESULT 42

MYHA_BOVIN STANDARD; PRT; 1976 AA.

ID MYHA_BOVIN STANDARD; PRT; 1976 AA.

AC 027991; 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Myosin heavy chain, nonmuscle type B (Cellular myosin heavy chain, type B) (Nonmuscle myosin heavy chain-B) (NMHC-B).

GN MYH10.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.

OC NCBI_Taxid=9913;

OX [1]

RN SEQUENCE FROM N.A.

RA Ohara M., Ishiguro N., Shingawa M.;

RT "Bos taurus nonmuscle myosin heavy chain B mRNA, complete cds.";

RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

RN (2)

RP SEQUENCE OF 204-302 FROM N.A.

RC TISSUE=Brain cortex;

RX MEDLINE=95301542; PubMed=7782316;

RA Itoh K., Adelstein R.S.;

RT "Neuronal cell expression of inserted isoforms of vertebrate nonmuscle myosin heavy chain II-B.";

RL J. Biol. Chem. 270:14533-14540(1995).

CC - FUNCTION: CELLULAR MYOSIN APPEARS TO PLAY A ROLE IN CYTOKINESIS, CELL SHAPE, AND SPECIALIZED FUNCTIONS SUCH AS SECRETION AND CAPING (BY SIMILARITY).

CC - SUBUNIT: MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2 HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC) AND 2 REGULATOR LIGHT CHAIN SUBUNITS (MLC-2).

CC - DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES, CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.

CC - SIMILARITY: Contains 1 myosin-like globular head domain.

CC - SIMILARITY: Contains 1 IQ domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AB022023; BAA36494.1; -

DR EMBL; U15716; AAB87715.1; -

DR HSSP; P10587; 1BR2.

DR InterPro; IPR000048; IQ_region.

DR InterPro; IPR001609; myosin_head.

DR InterPro; IPR004009; Myosin_N.

DR InterPro; IPR002928; Myosin_tail.

DR InterPro; IPR002017; Spectrin.

DR Pfam; PF00612; IQ; 1.

DR Pfam; PF00063; myosin_head; 1.

DR Pfam; PF02736; myosin_N; 1.

DR Pfam; PF01576; Myosin_tail; 1.

DR PRINTS; PR00193; MYOSINHEAVY.

DR PRODOM; PD000355; myosin_head; 1.

DR SMART; SM00015; IQ; 1.

DR SMART; SM00242; MYSC; 1.

DR PROSITE; PS50096; IQ; 1.

DR Myosin; ATP-binding; Calmodulin-binding; Actin-binding;

KW Coiled coil; Alkylation; Multigene family.

KW MYOSIN HEAD-LIKE.

FT DOMAIN 1 785

FT DOMAIN 786 815

FT DOMAIN 845 1976

FT NP_BIND 178 185

FT MOD_RES 701 701

FT MOD_RES 711 711

FT SEQUENCE 1976 AA; 229097 MW; 6144354451COF790 CRC64;

Query Match 3.3%; Score 135.5; DB 1; Length 1976;

Best local Similarity 18.5%; Pred. No. 18;

Matches 152; Conservative 120; Mismatches 335; Indels 213; Gaps 30;

QY 13 KENNRSYIDGQATOKTENLPDEVSKEGINAEQIVIKITDQGYTSHGDH---YHYH 69

DB EERNQILQNEKKKMQAHIQDL-EEQLDEEGARQKQLKBAKIKKMEIEILLLEQD 993

QY 70 NGKVPDAIISELLMKDPNYKLBEDIYNEVKGGVIVDGGKYTYLLDAADANVRK 129

DB 994 NKR-----PIKKKKMEDRIACSSQLAEERKAKNLAKIRKQEWISD---LEERLKK 1045

QY 130 EEINROKQESQHR-EGGTPRNDGAVALLARSGRYTTDDGYTFNMSDIIEDTGDATVPH 188

DB 1046 EEKTRQLEKAKRKLDGETTDDQDIAELQAO---IDLKIQVAKKEELQG---ALAR 1098

QY 189 GDHYHYIPRX-----ELASGLAAAEAFSLSGRGLNSRTYRONS----- 229

DB 1099 GDD-ETLHKNNALKVYRELQAOIAELQEDFEESEKASRNKAQKRDLSLEALKTLELD 1157

QY 230 --DNISRTWVSVPSTGTTNTNTSNTNSQASQND-----ISLKLQYKLP 279

DB 1158 TDDTTAAQOELRTKEQEVAAELKKALEERTKHEAQIQMRQRHATALELSEQLQAR 1217

QY 280 SQRHVESD--GLVFPDAQITSTRAGVAV-PHQDHYHFIYSGMSELEERAR---IIP 332

DB 1218 FRANLEKXKQGLTQNKELACEVKYLOQVKAESHRKKLDQOVOLHAKVSGDRLRVE 1277

QY 333 LFRSNHWVPDSRPEQSPQPTPEPSPGQPAENLKIDSSNLSVQVYKVGEGYFEER 392

DB 1278 LAEKNK-----KLQNELNVSTLLSEA-----EKK 1302

QY 393 GISRIFYAKDLPSEYKYNLESLKQESVSHLTAKENVAARDQEFYKAYNLLTEARK 452

DB 1303 GIK--FAND-----AAGLESQLODTQELQEBETQKMLSSIRQLSEERSLOQOE 1354

QY 453 ALFXNKGNSDFQALDKLERLNDESTNKEKLVDDLAFATITPERGKNSQIEYTE 512

DB 1355 ---EEPARS---LEKQALQAOQTTDYKKVDDDLGTENLEAKK--KLLKDEVUS 1405

QY 513 DEVRIAGLA-DKYTTSQGYTPE-HDIISDEGDAVTPMGHSHWIGKSLSDKEKVAQ 570

DB 1406 QLEEKALAYDLKELTKRTLQOELDLVD-----LDHQROQ-VSNLEKKQKKFPQ 1455

QY 571 AYTKEXGILPESPDAVDKANPTGDSAAIYNRVKGEKRIPLVRLPYMEHTVEVKNGLI 630

DB 1456 LLAEEKNI-----SARYAEERDRAEA-EAREKETKALSLAR----- 1490

QY 631 IHHKHVHNIRKFAWFDHTYKAPNGVLTLEDLFAITKYVYEHDERHSDGMSGEHYL 690

DB 1491 ----- 1496

QY 691 GKHDSEDPNKNFKADEEP-----VETPAEPVPOVETEKVEAOKEAEVLVA 739

DB 1497 EAREEAEERKQRLRADMEDLMSSKDDVGKNGVHELEKSKALAEQVEEMTQLBELEDEIQ 1556

QY 740 KYTDSLKANATETLAGLRNNLTLOIMDNNSIMAEAEKLL 779

DB 1557 ATEDAKRLLEVNMQAMKAQFERDLQTRDEQN--EEKRILL 1594

RESULT 43

PGCV_MOUSE STANDARD; PRT; 3358 AA.

ID PGCV_MOUSE STANDARD; PRT; 3358 AA.

AC Q62059; Q62058; Q9CUD0;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Versican core protein precursor (large fibroblast proteoglycan) (Chondroitin sulfate proteoglycan core protein 2) (PG-M).

GN CSPG2.

OS Mus musculus (Mouse).

CC	-1- SIMILARITY: Contains 1 C-type lectin family domain.
CC	-1- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC	-1- SIMILARITY: BELONGS TO THE AGRFCAN/VERSICAN PROTEOGLYCAN FAMILY.
CC	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL, D16263, BAA03796.1, -
DR	EMBL, D28599, -, NOT_ANNOTATED_CDS.
DR	EMBL, D32040, BAA06802.1, -
DR	EMBL, AK014525, BAB99411.1, -
DR	HSSP, P01132, IEPG.
DR	MCD, MGI:102889, Cspg2.
DR	InterPro, IPR000152, Asx_hydroxyl.
DR	InterPro, IPR000742, EGF_2.
DR	InterPro, IPR001881, EGF_Ca.
DR	InterPro, IPR006209, EGF_Like.
DR	InterPro, IPR007110, Ig-Like.
DR	InterPro, IPR003599, Ig.
DR	InterPro, IPR003006, Ig_MHC.
DR	InterPro, IPR001304, Lectin_C.
DR	InterPro, IPR000538, Link.
DR	InterPro, IPR000436, Sushi_SCR_CCP.
DR	Pfam, PF00008, EGF_2.
DR	Pfam, PF00047, Ig_1.
DR	Pfam, PF00059, lectin_C_1.
DR	Pfam, PF00084, sushi_1.
DR	Pfam, PF00193, Xlink_2.
DR	PRINTS, PF01265, LINKMODULE.
DR	ProDom, PD000918, Ccp_1_2.
DR	SMART, SM00032, Ccp_1.
DR	SMART, SM00034, CLECT_1.
DR	SMART, SM00179, EGF_CA_1.
DR	SMART, SM00409, Ig_1.
DR	SMART, SM00445, Link_2.
DR	PROSITE, PS00010, ASX_HYDROXYL_1.
DR	PROSITE, PS00615, C_TYPE_LECTIN_1_1.
DR	PROSITE, PS00041, C_TYPE_LECTIN_2_1.
DR	PROSITE, PS00022, EGF_1_2.
DR	PROSITE, PS01186, EGF_2_1.
DR	PROSITE, PS01187, EGF_CA_1.
DR	PROSITE, PS050835, IG_Like_1.
DR	PROSITE, PS01241, Link_2.
KW	Glycoprotein, Proteoglycan, Lectin, Extracellular matrix, Sushi;
KW	Signal, Repeat, EGF-like domain, Calcium, Immunoglobulin domain;
KW	Hyaluronic acid, Alternative splicing.
FT	SIGNAL 1 20
FT	CHAIN 21 3358
FT	DOMAIN 21 146
FT	DOMAIN 167 244
FT	DOMAIN 265 346
FT	DOMAIN 348 1308
FT	
FT	DOMAIN 1309 3052
FT	DOMAIN 3052 3088
FT	DOMAIN 3090 3126
FT	DOMAIN 3139 3253
FT	DOMAIN 3258 3316
FT	DISULFID 44 130
FT	DISULFID 172 243
FT	DISULFID 196 217
FT	DISULFID 270 333
FT	DISULFID 294 315
FT	DISULFID 3056 3067
FT	DISULFID 3061 3076
FT	DISULFID 3078 3087
FT	DISULFID 3094 3105
FT	DISULFID 3099 3114
FT	DISULFID 3114
FT	
FT	DOMAIN 1309 3052
FT	DOMAIN 3052 3088
FT	DOMAIN 3090 3126
FT	DOMAIN 3139 3253
FT	DOMAIN 3258 3316
FT	DISULFID 44 130
FT	DISULFID 172 243
FT	DISULFID 196 217
FT	DISULFID 270 333
FT	DISULFID 294 315
FT	DISULFID 3056 3067
FT	DISULFID 3061 3076
FT	DISULFID 3078 3087
FT	DISULFID 3094 3105
FT	DISULFID 3099 3114
FT	DISULFID 3114
FT	
FT	DOMAIN 1309 3052
FT	DOMAIN 3052 3088
FT	DOMAIN 3090 3126
FT	DOMAIN 3139 3253
FT	DOMAIN 3258 3316
FT	DISULFID 44 130
FT	DISULFID 172 243
FT	DISULFID 196 217
FT	DISULFID 270 333
FT	DISULFID 294 315
FT	DISULFID 3056 3067
FT	DISULFID 3061 3076
FT	DISULFID 3078 3087
FT	DISULFID 3094 3105
FT	DISULFID 3099 3114
FT	DISULFID 3114
FT	
FT	DOMAIN 1309 3052
FT	DOMAIN 3052 3088
FT	DOMAIN 3090 3126
FT	DOMAIN 3139 3253
FT	DOMAIN 3258 3316
FT	DISULFID 44 130
FT	DISULFID 172 243
FT	DISULFID 196 217
FT	DISULFID 270 333
FT	DISULFID 294 315
FT	DISULFID 3056 3067
FT	DISULFID 3061 3076
FT	DISULFID 3078 3087
FT	DISULFID 3094 3105
FT	DISULFID 3099 3114
FT	DISULFID 3114
FT	
FT	DOMAIN 1309 3052
FT	DOMAIN 3052 3088
FT	DOMAIN 3090 3126
FT	DOMAIN 3139 3253
FT	DOMAIN 3258 3316
FT	DISULFID 44 130
FT	DISULFID 172 243
FT	DISULFID 196 217
FT	DISULFID 270 333
FT	DISULFID 294 315
FT	DISULFID 3056 3067
FT	DISULFID 306

```

FT DISULFID 3116 3125 BY SIMILARITY.
FT DISULFID 3132 3143 BY SIMILARITY.
FT DISULFID 3160 3252 BY SIMILARITY.
FT DISULFID 3228 3244 BY SIMILARITY.
FT DISULFID 3259 3302 BY SIMILARITY.
FT DISULFID 3288 3315 BY SIMILARITY.
FT CARBOHYD 57 57 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 351 351 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 441 441 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 807 807 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 914 914 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 951 951 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1305 1305 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1372 1372 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1679 1679 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2054 2054 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2244 2244 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2362 2362 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2627 2627 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3030 3030 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3332 3332 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 3342 3342 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARSPLIC 348 348 P -> R (in isoform VI and isoform V3).
FT VARSPLIC 349 1308 /FTId=VSP 003087.
FT VARSPLIC 1309 3052 Missing (in isoform V1).
FT VARSPLIC 349 3052 /FTId=VSP 003088.
FT VARSPLIC 349 3052 Missing (in isoform V2).
FT VARSPLIC 349 3052 /FTId=VSP 003089.
FT VARSPLIC 349 3052 Missing (in isoform V3).
FT CONFLICT 126 126 A -> G (IN REF. 3).
FT CONFLICT 348 348 /FTId=VSP 003090.
FT CONFLICT 1658 1658 I -> T (IN REF. 3).
FT CONFLICT 1674 1680 TWNNSS -> QPQIGTA (IN REF. 3).
SQ SEQUENCE 3358 AA; 366938 MW; 071B80026BC0762D CRC64;

Query Match 3.3%; Score 135.5; DB 1; Length 3358;
Best Local Similarity 20.1%; Pred. No. 37;
Matches 143; Conservative 88; Mismatches 273; Indels 209; Gaps 34;

QY 239 PSVSNPGTNTNTNSNTNSQASQ-SNDIDSLKLYKPLSQRHVESDGLVFPQAQT 297
DB 2346 PTHRPQTMGSLHNSNSASASEBEGATPTAFLPQTVSEVMTKHPAPSESQSDLFVNV 2405
QY 298 STTAG-----VAVPHD-----HYHPIPSQMSLEERARIPL----- 333
DB 2406 SGEGSEVDTLDLVYTSGTTOASSQDSMLASHGFLKHPKPEVSKTEGATDVSPTASAMF 2465
QY 334 ----RYRSHWYVDSRPEQSPQRPQAPNKLKIDNSSLSVQLVHKVGGVVF 389
DB 2466 LHMSEYKSLY-PTS--TLPSSTPYKSPBEGIDGQDNIOFEGSTLAKSRRTTESIT 2522
QY 390 E-----EKG--ISRYVPAKDLPESETV-KNL-----ESKLS 416
DB 2523 DLDKEDSKDLGLTTTESAIVKSLPELTSKNIIIDHTRKPYEYIPIQIOTLDPEIKLE 2582
QY 417 KQESVSHITLAKKE-----NVAPEQDEFYDKAVNLTETAKLFX-----NKGNSDFQA 466
DB 2583 SHGSSEESIQQVQEKYEGAVTLSPTEBSPGSGDALAGYTOAIVNESVTPNDGK---QA 2638
QY 467 LD---KLRLNDESTNKEKLVDDLAFL--APITH-PERLCKPNQIEYTDDE-VRIMQ 519
DB 2639 EDISSFATGIPVSTETE---LHTFPPTASTLIPKLTATASPEIDKPNIEATSLND 2693
QY 530 IADKXTTSDGY-IPEHDIISDEGDAYVTPHMGSHMICKDS-----LSD--- 563
DB 2694 IFESSTLSGGAIADQSEVISTGLHLEKQOEYEEKYGGPSPQPFSGVGEVLDPDA 2753
QY 564 -----KEKAAQAYTEKGLIPSPPAD-VKANPTGDSAAAIYNRVKGKRIPLVRLPM 617
DB 2754 VVISGISTYIAQTLTELPMVVRPDSSTHYTEATPEVSSIAEL-----SPOIPSPSPFP-- 2805

```

```

QY 618 VEHTEVKNGLIIEPKDHYNIKFAMFDDHTYKANGYTLDELPAI---KYVEHPDE 674
DB 2806 ----YVYDNGVSKFPEVPH-----TSAQPVSTYSSQKSIESPCK 2841
QY 675 RHNSDGNMASEHV--LGKDH-----SDPKNKFKAD-----EPVETPAEPV 719
DB 2842 EVHA-----NIEETIKPLGVNHRTEPPSMRDPALDVSEDESKHLLEELTSPKPEP 2896
QY 720 PQ-----VETEKVQAOLKEADEVLLAKYTDSSILKANAT----- 751
DB 2897 SODPNNKAKDHPGTVGMAGIRTTSEPVITADMELGATQOPHSGSAAFVETGMV 2956
QY 752 -----ETLAGLR-NMLTLQIMDNNSIWAEBKXLA--LLKGNPSSVS 791
DB 2957 PQIQGEPERPFPFLEINHEHTLSLPFESILATSEKQVSQILNNSQATVS 3009

RESULT 44
ID IGA_NEIGO STANDARD; PRT; 1532 AA.
AC P09790;
DT 01-MAR-1989 (rel. 10, Created)
DT 01-MAR-1989 (rel. 10, Last sequence update)
DT 28-FEB-2003 (rel. 41, Last annotation update)
DE IGA-specific serine endopeptidase precursor (EC 3.4.21.72) (IGA
DE protease).
GN IGA.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales;
OC Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=MS11;
RC MEDLINE=87115823; PubMed=3027577;
RA Pohner J., Halter R., Beyreuther K., Meyer T.F.;
RT "Gene structure and extracellular secretion of Neisseria gonorrhoeae
RT IGA protease.";
RL Nature 325:458-462 (1987).
RN [2]
RP ACTIVE SITE.
RX MEDLINE=90154052; PubMed=2105953;
RA Bachovichin W.W., Plaut A.G., Plentke G.R., Lynch M., Kettner C.A.;
RT "Inhibition of IGA1 proteinases from Neisseria gonorrhoeae and
RT Hemophilus influenzae by peptide prolyl boronic acids.";
RL J. Biol. Chem. 265:3738-3743 (1990).
CC -I- FUNCTION: THIS PROTEASE IS SPECIFIC FOR IMMUNOGLOBULIN A.
CC -I- CATALYTIC ACTIVITY: Cleavage of immunoglobulin A molecules at
CC certain Pro-I-Xaa bonds in the hinge region. No small molecule
CC substrates are known.
CC -I- SUBCELLULAR LOCATION: Secreted.
CC -I- DOMAIN: THE SIGNAL PEPTIDE GUIDE THE PRECURSOR TO THE PERIPLASMIC
CC SPACE, AND THE CARBOXY-TERMINAL HELPER DOMAIN ASSOCIATES WITH THE
CC OUTER MEMBRANE TO FORM A PORE FOR EXCRETION OF THE PROTEASE
CC DOMAIN. THE HELPER DOMAIN IS THEN RELEASED BY AUTOPROTEOLYSIS.
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S6.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X04835; CAA28538.1; -.
CC PIR: A26039; A26039.
CC MEROPS: S06.001; -.
CC InterPro: IPR006315; Autotransport.
CC InterPro: IPR005546; Autotransporter.
CC InterPro: IPR000710; IGA_S6.
CC InterPro: IPR004899; Pertactin.
CC Pfam: PF03797; Autotransporter; 1.

```

QY	7	YQARIVKKNRNRYSYDGOATQKTENLT--PDEVSKREGI--NAEQIVIKITDQGYTSHD	64
Db	731	EKAAEIAANQASFSFGSRNVSDITANITATDPAKNIKLGKNGDEVCVASDITYGTCN--	788
QY	65	HYHYNGKVPYDAIISEELMKDPNYLKDEDI VNEVKGVYIKVDKXYVYILKDAHAAD	124
Db	789	-----TGNLSDPAKNSFPDATTRINGVNINQNAALVIGKAAIMGKIQGQ-----GNSRVS	838
QY	125	NVRTKEELNRQKQEHSHQREGGTPRNDQAVL-----ARSGCRYYT-----DGYIFN	172
Db	839	NQHSKMHLTGDSQVHN-----LSLADSHIHLMNNSADQASANKYHTIKIMHLSGHPHY	892
QY	173	ASDIIEDGDYIVPH--GDHYHYIKPKNELASSELAAEAF-----LSGRCNLSNRG---	223
Db	893	LTDLAKNIGDKVLVESASGHQIAHVNQNTGEPNOEGDLPASSVODRSRLFYSLAHY	952
QY	224	-----YRQNSDNTSRRTNWVPYSNPGTNTNTNSNNTNSQAS-----	262
Db	953	VDLGLARTTIKTENITRL--YNPYANGNRPVYPARPSAANTASQAKATQTDQAQIARQ	101
QY	263	-----QSDIDISLLKOLYKPLPSQRH--VESDGLVDEPAQITSRTARGVAVPHGDH	311
Db	1012	NIWVAPSPQANQAEALRQAKAQVKQKQAAEAKVARKQDEBAKRAEIAIQOEBA	1073
QY	312	YHFIPIYSQMSLE-ERIRAIIPLRKRSNMHWVDSRPEQSPPTPEPS-----PQPQ	362
Db	1072	RKAABELAKQKAEARKARELARO-----KAEESHQANAKPKRRRRALILPRP-	1120
QY	363	PAPNLIKID-----SNSSL--VSOLQVKVGEYV--FPEKQISRYVFAKDLPSFETVKL	411
Db	1121	PAPVSLDDYDAKONSSESIGLAVIIPRMGELLINDYEE-----IPLSELED-	1166
QY	412	ESKLSQESVSHTLTAKKENAVAPRDOEFYDKAYNLLTEAHKALFYNKGRNSDFQALDKLL	471
Db	1169	EABEERRQAT-----QFHSKSRN-----RAIISEPSEDEDASSVS	120
QY	472	ERLNDGESTNKEKLVDDLAFAPRITHPERLGKPNQSEIYTEDEVR1--AQIADXYTSD	528
Db	1208	DKHPQDNTLHFKVE--TAGLQRP-----RAQPRTOAAQADAVSTNTWSALSDMASTQ	1266
QY	529	GYIPEHNIISDEGAYVTPRMHGSHWIKGDSLSDEKEXAQAQATYKKEGILPPSPDADYK	588
Db	1261	SILDLT-----GAYLTRH-----IAQKRAAEK-----NSW	1288
QY	589	ANPTG-----DSAAIYNRVKGEKRIPLVRLPYWHEHTVEYKNGNLIIPHKHQYHIKFAWF	645
Db	1289	MSNTQGYGDYASQYRRPFSKKTQTOQIGDRKLSENMQGQ---VLTYSDSQHT-----F	1344
QY	646	DDHTYKAENGYYLLEDFATIKXYVEHPDERPHNSDQGWGNASEHVLGK--KDHSEBENK--NF	703
Db	1341	DQAGSK--NTFYQANLYG--KYYL-----NDAMVYAGDIGAGSLRSRLQTOQKXNF	138
QY	704	KADEPVEETPAEPVPOVETEKVE-----AQKAEVLLAKVTDSILKAN--	749

ID	YMX6_YEAST	STANDARD	PRT	960 AA.
AC	004279			
DT	01-NOV-1997	(Rel. 35, Created)		
DT	01-NOV-1997	(Rel. 35, Last sequence update)		
DT	15-SEP-2003	(Rel. 42, Last annotation update)		
DE	Hypothetical 105.9 kDa protein in ADH3-RCAl intergenic region.			
CN	YMR086W OR YW9582.10.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
OX	NCBI_TaxID=4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=SS288C / AB972;			
RX	Pubmed-9163872;			
RA	Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,			
RA	Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,			
RA	Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,			
RA	Rice P., Skelton J., Walsh S., Whitehead S., Barrall B.G.;			
RT	"The nucleotide sequence of Saccharomyces cerevisiae chromosome			
RL	XIII.", 387:90-93(1997).			
CC	Nature			
CC	-I- SIMILARITY: SOME, TO YEAST YKL105C.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.1db-sib.ch/announce/			
CC	or send an email to license@1db-sib.ch).			
CC	-----			
DR	EMBL; 249259; CAA89232.1; -			
DR	PIR; S54461; S54461.			
DR	SGD; S0004692; YMR086W.			
KM	Hypothetical protein.			
SO	SEQUENCE 960 AA; 105873 MW; 05A4FA27129DB098 CRC64;			
	Query Match	3.2%; Score 134.5; DB 1; Length 960;		
	Best Local Similarity	21.8%; Pred. No. 7.4;		
	Matches 143; Conservative 87; Mismatches 249; Indels 178; Gaps 35.			
OY	196 PKNELASLSLAABA---FLISGGLNSNF---TYRQNSDN-----TSRTWVPEVS 242			
DB	9 PEMEOADPTAAVAASIGKLLMKKGKQNSDNEQRPETYSASMTNLRKPSAPKRMSSISS 68			
OY	243 NPGTTNTNTSNNSTNS--QASQSDNDLSLLKOLYKLP--LSORHVESDGLVFPDAQTS 298			
DB	69 SESRRSDGKRPGLKINSITLQSSMGKEDLSLPTKEPQHTTRSHNRTSLSPNQRQ-QS 127			
OY	299 RTAGVAVPHGDHYHPIYSQMSLEERIAHPIPLRYRSNHWVDSRPEQSPQPTPEPS 358			
DB	128 RNSGSL-----QRKSKTHORISYDEAKQTFPKDGGPGQAGILTGQARTENPS 175			
OY	359 -----PPQPARPNKI-----DSNSLSVQLVRKKGDEY-VPEEKGISRYVAK 401			
DB	176 GSIPLRTTKRYIPGNGVLAIVEKEPSNAN-TSKLIRSNSSAHSLMANNGS--LIRK 232			
OY	402 DLPEETVANKLESKSKQESVSHLTAKKENAVAPRDOEYDKAVNLLTEAHKALFXNKGKN 461			
DB	233 KVGQESLHSQPKTKSSLGNTSS-T-QAKKCGKAVQERN-----LAKKHDI-----N 276			
OY	462 SDF-----QALDKLIERLNDESTNKEKLVYDLLAFIAFTPTPERLGRKNSQIETVED 513			

```

Db 277 SNVPLIEVREBETDDELKLDNSNSESESTVNS-----ENNLEKPSG-LINIEKD 325
Qy 514 EVRIALQADKYTTSDGYIFD---EHDIIIDEGDAYVTPH--MGSHWIGKDSLDEKXVA 568
Db 326 D--LSKLIHENTLEBSEFIEKKEKRLNDQBEVTFSEKTVKGDVW-----PKXN 374
Qy 569 AQAAYTEK-----GIIP---PSPDADVKANPTGDSAAAIYKVGKGRIPLYR 613
Db 375 RQASTLEKTFNYDNEKEKRGKPVABHPLEAEV-----DDKSVGEQNNISSSG----- 423
Qy 614 LPMYMEHYE---VKNGNLI-----PHKDHNIKAMPDHDHYKAPNGTLEDLRA 663
Db 424 -SYSAQGSYENSKPEKNGKLENDTSSPTQD-----LDEKRSIEKDTLEGTES 473
Qy 664 TIKYVYEHDPDRPHNSDGNASEHYLGKKDSEDPKNKPKADE--EPVEETPAEPVQ 721
Db 474 SKK--VEKPD-----NCCERKISGANTSSKKKNVGDYKDEFDYTES-----D 514
Qy 722 VETKEVQAQKEAEVLAKVTDLSLKANATETIAG-LRNNLTLOIMDNNSIMAAEAK 777
Db 515 QKTSKSN-----SKNTDHSNOTBPTPLSLAQVLRFSNTYLSKKNQSKQAEDEK 562

RESULT 46
TCF8_HUMAN STANDARD; PRT; 1124 AA.
AC P37275; Q12924; Q13800;
ID 01-OCT-1994 (Rel. 30, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transcription factor 8 (N1L-2-A zinc finger protein) (Negative
DE regulator of IL2).
OS Homo sapiens (Human).
GN TCF8 OR AREB6.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP MEDLINE=94186507; PubMed=8138542;
RX Watanabe Y., Kawakami K., Hirayama Y., Nagano K.;
RT "Transcription factors positively and negatively regulating the Na,K-
RT ATPase alpha 1 subunit gene.";
RL J. Biochem. 114:849-855(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Bachman N.J., Scarputta R.C.;
RT "A human zinc finger homeodomain protein homologous to the chicken
RT delta-crystallin enhancer binding protein, delta E1.";
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 390-1124 FROM N.A.
RX MEDLINE=92108424; PubMed=1840704;
RA Williams T.M., Moolten D., Burtin J., Romano J., Bhaerman R.,
RA Goddard A., Mellon M., Rauscher F.J. III, Kant J.A.;
RT "Identification of a zinc finger protein that inhibits IL-2 gene
RT expression.";
RL Science 254:1791-1794(1991).
CC -I- FUNCTION: INHIBITS INTERLEUKIN-2 (IL-2) GENE EXPRESSION. MAY BE
CC RESPONSIBLE FOR TRANSCRIPTIONAL REPRESSION OF THE IL-2 GENE.
CC ENHANCES OR REPRESSES THE PROMOTER ACTIVITY OF THE IL-2 GENE
CC DEPENDUNG ON THE QUANTITY OF CDNA AND ON THE CELL TYPE.
CC -I- TISSUE SPECIFICITY: EXPRESSED IN HEART AND SKELETAL MUSCLE, BUT
CC NOT IN LIVER, SPLEEN, OR PANCREAS.
CC -I- SIMILARITY: BELONGS TO DELTA-E1/2FH-1 FAMILY OF TWO-HANDED ZINC
CC FINGER/HOMEODOMAIN PROTEINS.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D15050; BAA03646.1; -
DR EMBL; U12170; AAA20602.1; -
DR EMBL; M81699; -; NOT ANNOTATED_CDS.
DR PIR; JX0293; JX0293.
DR TRANSFAC; T00625; -
DR Genew; HGNC:11642; TCF8.
DR MIM; 189909; -
DR GO; GO:0003713; F:transcription co-activator activity; TAS.
DR GO; GO:0003714; F:transcription co-repressor activity; TAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0008270; F:zinc ion binding activity; TAS.
DR GO; GO:0008283; P:cell proliferation; TAS.
DR GO; GO:0005955; P:immune response; TAS.
DR GO; GO:0001224; P:negative regulation of transcription from P. .; TAS.
DR InterPro; IPR001356; Homeobox.
DR Pfam; PF00096; zf-C2H2; 7.
DR ProDom; PD000010; Homeobox; 1.
DR ProDom; PD000003; Znf_C2H2; 2.
DR SMART; SM00389; HOX; 1.
DR SMART; SM00355; Znf_C2H2; 7.
DR PROSITE; PS00028; ZINC FINGER C2H2_1; 5.
DR PROSITE; PS50157; ZINC FINGER C2H2_2; 6.
KW Transcription regulation; DNA-binding; Nuclear protein; zinc-finger;
KW Homeobox; Repressor; Activator; Metal-binding; Repeat.
FT ZN_FING 170 193 C2H2-TYPE.
FT ZN_FING 200 222 C2H2-TYPE.
FT ZN_FING 240 262 C2H2-TYPE.
FT ZN_FING 268 292 C2H2-TYPE (ATYPICAL).
FT DNF_BIND 581 640 HOMEBOX-LIKE.
FT ZN_FING 904 926 C2H2-TYPE.
FT ZN_FING 932 954 C2H2-TYPE (ATYPICAL).
FT ZN_FING 960 981 C2H2-TYPE (ACIDIC).
FT DOMAIN 989 1124 GLU-RICH (ACIDIC).
FT CONFLICT 420 420 V -> I (IN REF. 2).
FT CONFLICT 609 609 E -> Q (IN REF. 3).
FT CONFLICT 654 654 I -> T (IN REF. 2).
FT CONFLICT 672 672 D -> H (IN REF. 3).
FT CONFLICT 681 681 L -> S (IN REF. 3).
SQ SEQUENCE 1124 AA; 124073 MW; 0A2714CC37C848D1 CRC64;

Query Match 3.2%; Score 134.5; DB 1; Length 1124;
Best Local Similarity 19.0%; Pred. No. 9.2;
Matches 168; Conservative 104; Mismatches 323; Indels 289; Gaps 34;

Qy 21 IGGKATOKTEMLTPDEVSK-REGINAEQIVIKITDQGVTSHGPHYHYNGKVPYDAII 79
Db 420 VDNVIRQVLENNQANLASKQETINASF-----OQGH-----SVI 457
Qy 80 SEELMKDPNYKLKDEDIYNEKGVYIKVDGKYVYLKDAAHADNVRTKE-----EI 132
Db 458 SAISL-----PLVQDGTTKIINYSLEQPSQLGVPPNKKENKPVANNSCKSEKLPD 511
Qy 133 NFQKQEHSGRGGGPRNDGAVALARSGQRYTTDQGYIFNASDIETDGDATIVPHGHY 192
Db 512 LTVKSEKSKSPFGVY--NDSTCLC-----DD-----CPDINALDELKAY 550
Qy 193 HTIPKNEIASASIAAEA-----PLSGRGNLSNRYRRQNSDNTSRTMWSVSNP 244
Db 551 DL--KQTPPPPLPAHEKEKPESSVSATGDCNLSPOPLK-----NLSTILKAY 599
Qy 245 GTTNTNTSNTNSNTNSQASQNDISLLKQLYKLPLSQRHVESD-----GLVFPQAQT 297
Db 600 YALNQPSAEELSKTADSVNLPDVKWKFEMQAGQISVQSSSESPSPGKVNIPAKNN 659
Qy 298 SRTANGAVPHGDHYHFLPYQMSLEERIAIILPLRIRSNHWVPDSRPOSPQPTPP 357
Db 660 DQPOSANNE-----FOSTVNLQ-----PLKMTNSPVLVPGVSTTNSRSSTSP 705
Qy 358 SPGPQAPNLIKIDNSLSLVQVRAVGSGYVFEKIGISRYPAK-----DLPEYKMLE- 412

```

```

Db 706 SP-----LNTSSRRNT-----CGYLTAAGAGQEPVEPDLSTLFPQOCELLER 749
Oy 413 -----SKKSQESVSHLTAKKE-----NVARDFEYKANKNLTE 449
Db 750 STTSVONSVYSVVEEPLNLSCKAKEPKQKSCVTDESVVNVIPPSANPINALPTVA 809
Oy 450 AHKALFXNKGNSDQALDKLERLNDESTNKEKLVLDLAFAPIT-----HPERLG 502
Db 810 QLPFTVALADONS-----VPCRLALANKOTLIPVATYTTSTYSPAVQEPFLAVI 861
Oy 503 KPNQOIEYTEDEVRIALQADKTTSDGYIFDEHDIISBEGDAYVTPHNGSHWICKDLS 562
Db 862 QPNQOBEROD-----TSSEGV-----SNVEDQNDSDSP----- 891
Oy 563 DKEKVAQAAYTEKELIPSPADVKANPTGSAALINRVGSEKRIPLVRLPYVVEHTV 622
Db 892 -----PKKKMKRTKENGMYACDLCDKI-FQKSSSLRRKY--EHTG 928
Oy 623 E-----VKNGNLIPH-----KDHYNIKFAMFDPDHTYKAPNG 655
Db 929 KRPHECGICCKAFKFKHHLIEHMLHSGEKPYQCKCKGRFSHSGSYQHMHR----- 983
Oy 656 YLLEDLFATIKYVHPDER-----PHSNDGNGASEHVLGKQDS 696
Db 984 -----SYCKREAEERDSTEOEAGPEILSNHVGARASFGQSDERESLTERED 1033
Oy 697 EDPNKNFPADEPVEVETAEPEV--PQ--VETEKYEAQLEKENVLAKYDSSLKANATE 752
Db 1034 EDSEKEEEDDEKEMELOEKECEKPKQDEEEEEEVEEVEEVEANEGEAK----- 1088
Oy 753 TLAGRNLTLOIMNNSIMAEAKTLALTKGNSPSSYSKEKIN 796
Db 1089 -----TEGLMKDDRAESQASS-LQKVESSESGVSEKTN 1122

RESULT 47
RTN4 RAT STANDARD: PRT: 1163 AA.
ID RTN4 RAT
AC Q9JKL1; Q9JKL10; Q9ROD9; Q9WUE9; Q9WUF0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Reticulon 4 (Neurite outgrowth inhibitor) (Nogo protein) (Foocen)
GN (Glut4 vesicle 20 kDa protein).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 3), AND PARTIAL SEQUENCE.
RC STPAIN:Sprague-Dawley; TISSUE=adipocyte;
RX MEDLINE=99249816; PubMed=10231557;
RA Morris N.J., Ross S.A., Neveu J.M., Lane W.S., Lienhard G.E.;
RT "Cloning and characterization of a 22 kDa protein from rat adipocytes:
RL a new member of the reticulon family.";
RL Biochim. Biophys. Acta 1450:68-76(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RX MEDLINE=20129258; PubMed=1067796;
RA Chen M.S., Huber A.B., Van der Haar M.E., Frank M., Schnell L.,
RA Spillmann A.A., Christ F., Schwab M.E.;
RT "Nogo-A is a myelin-associated neurite outgrowth inhibitor and an
RL antigen for monoclonal antibody IN-1.";
RL Nature 403:434-439(2000).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 4).
RC STPAIN:Wistar Kyoto; TISSUE=vascular smooth muscle;
RA Ito T., Schwartz S.M.;
RT "Cloning of a member of the reticulon gene family in rat: one of two
RT minor splice variants.";
RT Submitted (FEB-1999) to the EMBL/Genbank/DBJ databases.

```

```

RN [4]
RP FUNCTION.
RX MEDLINE=22033691; PubMed=12037567;
RA Grandpre T., Li S., Strittmatter S.M.;
RL "Nogo-66 receptor antagonist peptide promotes axonal regeneration.";
RL Nature 417:547-551(2002).
CC -1- FUNCTION: Potent neurite outgrowth inhibitor which may also help
CC block the regeneration of the nervous central system in adults (By
CC similarity).
CC -1- SUBUNIT: Binds to RTN4R. Interacts with Bcl-x1 and Bcl-2 (By
CC similarity).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Anchored to the
CC membrane of the endoplasmic reticulum through 2 putative
CC transmembrane domains (By similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1; Synonyms=Nogo-A, NI-220-250;
CC IsoId=Q9JUK1-1; Sequence=Displayed;
CC Name=2; Synonyms=Nogo-B, Foocen-M1;
CC IsoId=Q9JUK1-2; Sequence=VSP_005658;
CC Name=3; Synonyms=Nogo-C, VP20;
CC IsoId=Q9JUK1-3; Sequence=VSP_005656, VSP_005657;
CC Name=4; Synonyms=Foocen-M2;
CC IsoId=Q9JUK1-4; Sequence=VSP_005659;
CC -1- TISSUE SPECIFICITY: Isoforms 1, 2 and 3 are present in optic
CC nerve, spinal cord and cerebral cortex. Isoforms 1 and 2 are
CC present in dorsal root ganglion, sciatic nerve and PC12 cells
CC after longer exposure. Isoforms 2 and 3 are detected in kidney,
CC cartilage, skin, lung and spleen. Isoform 3 is expressed at high
CC level in skeletal muscle. In adult animals isoform 1 is expressed
CC mainly in the nervous system.
CC -1- SIMILARITY: Contains 1 reticulon domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF051335; AAF01564.1; -
DR EMBL; AJ242961; CAB71027.1; -
DR EMBL; AJ242962; CAB71028.1; -
DR EMBL; AJ242963; CAB71029.1; -
DR EMBL; AF132045; AAB31019.1; -
DR EMBL; AF132046; AAB31020.1; -
DR GO; GO:0030176; C:endoplasmic reticulum membrane, intrinsic p...; IDA.
DR GO; GO:0005635; C:nuclear membrane; ISS.
DR GO; GO:0005515; F:protein binding activity; ISS.
DR GO; GO:0019987; P:negative regulation of anti-apoptosis; ISS.
DR GO; GO:0030517; P:negative regulation of axon extension; ISS.
DR InterPro; IPR003388; Reticulon.
DR Pfam; PF02453; Reticulon; 1.
DR PROSITE; PS50845; RETICULON; 1.
KW Endoplasmic reticulum; Alternative splicing; Transmembrane.
FT DOMAIN 1 989
FT TRANSHEM 990 1010
FT DOMAIN 1011 1104
FT TRANSHEM 1105 1125
FT DOMAIN 1126 1163
FT DOMAIN 976 1163
FT DOMAIN 33 46
FT DOMAIN 73 76
FT DOMAIN 140 145
FT TRANSPLIC 1 964
FT VARSPPLIC 965 975
FT VARSPPLIC 173 975
FT VARSPPLIC 192 975

```

```

FT CONFLICT 1130 1131 /FTID=VSE_005659.
SQ SEQUENCE 1163 AA; 126386 MM; 8CB894B09E94F0B6 CRC64;
Query Match 3.2%; Score 134; DB 1; Length 1163;
Best Local Similarity 21.2%; Pred. No. 10;
Matches 170; Conservative 117; Mismatches 310; Indels 204; Gaps 43;

OY 47 EGVIVKITD--GGVYVSHGDHYHYNGKAVYDAIISSELMKDPNRYKLDKEDIVNEVKGKY 105
DB 309 EGVIVRSKDKEDLVCSAALH-----SQESEPVGKEDVVSBE--KTMDFINEMQMSV 358
OY 106 VIKVDGKYYVY-----LKD-----AAADNVATKEINKQKQESHQREGTGP 148
DB 359 VAPVEEVADFKPPEQAMEVKDTYESSRVLARA-NVASK--VDRKCLDELSLEQK----- 411
OY 149 RNDGVALAR--SQGRYTTDDGYIFNNSDIIEPTGDAYIYPHGDHYHYIKNELISASELAA 207
DB 412 -----SLQKDEGR--MEDASFPTPEVVKDSRAVIT-----CASFTSATESTT 454
OY 208 AEAFLSGRNLNSRTYRRQNSDN-----TSRTNWPSVSNPDTTNTNNSNNSQMS 262
DB 455 ANTFLPELHTSENKTKDEKKIEBKAQITTEKTS--FKTSNPLV-----AV 499
OY 263 QSNIDISLKLQYKLPQSQRHVESDGLVDPQAQITSRTARGVAVPHGDHYHFIPIYSQMS 322
DB 500 QDSEAD-----YVTTDTL-----SKVTEAANSN--MPEGLPDLVQACESE 539
OY 323 LEEERARITPLRYRNSHWVPDSRPEQSPQPTPEPSPG-----POPA-PNLIKIDS--N 372
DB 540 INEAGTKIAVETKVD-LVQTESEALQESLYPTAQLPSEFEAEATPSPVLPDIMEAPLN 598
OY 373 SSVLQVLRKVGEGVFEKKGISRYFAADLPSETYKNLESLSKQESH--TLRAKKEN 431
DB 599 SLPS-----AGASVQPSVSPLEADPPVSYDSTIKLEPN 633
OY 432 VAPRDEPFYDKAYNLITFAHKALEFNKNG--RNSDFQALDKLERLDESTNKEKLVDDL 489
DB 634 PPP-----YEAMNV--ALKALGTKEGKKEPESFNA-----AVQETAPYISLACDLI 679
OY 490 AFLAPITPBERLQKPNQSIQIEYEDV-RIQLADKTYTSDGYIFEHDIISDEGAYVTP 548
DB 680 KETKLTSPSPDFNSYSEIAKEFKSVPEHAELVEDSPSS-----EPVDLFSDDSIPEVQ 735
OY 549 HHGHHWIGKDSLSDKEKVAQAAYTEKGIPEPSPA-----DVANPTGSGSA-A 598
DB 736 TQEEVVMIMKESILTEVSETVAO-HKEER--LCASPOLGKPYLESFQPMHSTKDAASND 792
OY 599 IYNRVKGEKRIPLVRLPYMVEHTVEYKGNLIIPHKDHVNIKFAWFDHTYKAPNGYTL 658
DB 793 IFTLTKKER-----ISIQMEEFNTAITSNDLSSKED--KIK-----ESETFSDDSPIEI 841
OY 659 EDLPATIKYYVHPDERPHSNDGWNASEHV-LGKKHSDSPNKPKADEPEVETPAEP 717
DB 842 IDEFPF---FVGAKDSDPKL-----AKGYTDLVSDKESEIANIGGASDLSCLPELPCDL 892
OY 718 EYPOV---ETEKVEAOLKE--AEVLLAKYTDSSIKANAETTLAAGLNNTLTQIMDNNGSM 772
DB 893 SFKNYTPKDEHVVSSEFSENRSVSXKASISPSNVSLALBPOTENG-----SIVXSKSLT 945
OY 773 AEAELKLLALLKGSNPSVSKE 793
DB 946 KEAEKKL-----PSDTEKE 959

RESULT 48
ID OSH1_YEAST STANDARD: PRT; 1188 AA.
AC P35845; P3555; P80234;
DT 01-UN-1994 (Rel. 29, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Oysterol-binding protein homolog 1.

```

```

GN OSH1 OR SWH1 OR YAR042W/YAR044W.
OC Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=95249563; PubMed=7731988;
RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA Hall J., Ouellette B.F.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA Storms R.K.;
RT "The nucleotide sequence of chromosome I from Saccharomyces
RT cerevisiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=DL1;
RA Schmalix W.A., Bandlow W.;
RL Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
RN (3)
RP CHARACTERIZATION.
RX MEDLINE=8017104; PubMed=8017104;
RA Jiang B., Brown J.L., Sheraton J., Fortin N., Bussey H.;
RT "A new family of yeast genes implicated in ergosterol synthesis is
RT related to the human oysterol binding protein.";
RL Yeast 10:341-353(1994).
RN (4)
RP SUBCELLULAR LOCATION.
RX MEDLINE=21301806; PubMed=11408574;
RA Levine T.P., Munro S.;
RT "Dual targeting of Oshp, a yeast homologue of oysterol-binding
RT protein, to both the Golgi and the nucleus-vacuole junction.";
RL Mol. Biol. Cell 12:1633-1644(2001).
RN (5)
RP GENETIC ANALYSIS.
RX MEDLINE=21135676; PubMed=1123839;
RA Beh C.T., Cool L., Phillips J., Rine J.;
RT "Overlapping functions of the yeast oysterol-binding protein
RT homologues.";
RL Genetics 157:1117-1140(2001).
CC -1- FUNCTION: PLAYS A ROLE IN ERGOSTEROL SYNTHESIS.
CC -1- SUBCELLULAR LOCATION: Found in Golgi and nucleus-vacuole (NV)
CC junction.
CC -1- SIMILARITY: BELONGS TO THE OSBP FAMILY.
CC -1- SIMILARITY: Contains 1 PH domain.
CC -1- SIMILARITY: Contains 3 ANK repeats.
CC -1- CAUTION: Ref.1 sequence differs from that shown due to a
CC frameshift in position 230.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL, L28920; AAC09497.1; ALT_FRAME.
DR EMBL, L28920; AAC09496.1; ALT_FRAME.
DR EMBL, X74552; CA52646.1; -.
DR SGD, S0000082; OSH1.
DR GO, GO:0005769; C:early endosome, IDA.
DR GO, GO:0000138; C:Golgi trans cisterna, IDA.
DR GO, GO:0005545; F:phosphatidylinositol binding activity, IDA.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR00648; Oysterol_BP.
DR InterPro: IPR001849; PH.
DR Pfam, PF00023; ank, 3.
DR Pfam, PF01237; Oysterol_BP, 1.
DR Pfam, PF00169; PH, 1.
DR SMART, SM00248; ANK, 2.
DR SMART, SM00233; PH, 1.

```

```

DR PROSITE; PS50088; ANK_REPEAT; 2.
DR PROSITE; PS50297; ANK_REPEAT_REGION; 2.
DR PROSITE; PS01013; OSBP; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
KM Lipid transport; Transport; Sterol biosynthesis; Repeat; ANK repeat.
FT REPEAT 51 80 ANK 1.
FT REPEAT 96 125 ANK 2.
FT REPEAT 196 225 ANK 3.
FT DOMAIN 330 379 PH.
FT DOMAIN 434 472 ASN/ASP-RICH.
FT DOMAIN 444 448 POLY-ASN.
FT DOMAIN 457 463 POLY-ASN.
FT DOMAIN 464 472 POLY-ASP.
FT CONFLICT 424 424 P -> A (IN REF. 2).
FT CONFLICT 495 496 TP -> AS (IN REF. 2).
SQ SEQUENCE 1188 AA; 135145 MW; 872952DFD2330207 CRC64;

Query Match 3.2%; Score 134; DB 1; Length 1188;
Best local similarity 18.7%; Pred. No. 11;
Matches 139; Conservative 103; Mismatches 256; Indels 244; Gaps 30;

41 REGINABQIVIKITDGGVTHSHGDHYHYNGKVPYDAIISSELT-----L 84
228 RKGLPIELVRKV-----NENYATNTKTAIDIELKLLERATBQSVIVDTNNL 278
85 MKDPNYKLDEDIVNEVGK---GVYIKVDGK--YVYIKDAADNVRTKEEINRQK-Q 137
279 HEAPTYKYVKKWTFAGQYKLRWFISSDGKLSYIIOADTKNA-----CRGSLMSGCS 334
138 EHSQHR-----GTPRNDGAVALARSOGRYTTDDGIENASDIIEDTGDAYIVPHGD 190
335 LHLDSSEKLKEEIIIG---NNGVIRPMHLKGNPIETNNWVAIQGAIRYAKDRILLANG 391
191 HY-----HYTPKELASSELAAEAFLSGKGNLSNR--TYRRQNSNTSRTNWVPSVS 242
392 PYSPLSHSGSSKVSXKKNLHAATSKRITPSHLSTLTQNDHDDDDSTNN-----N 446
243 NPGTNTNTNNSNTNSOASQNSNDISLLKOLYKPLSQRHVESDGLVDPQAITSRTAR 302
447 NKNKNDVDDNNNNNNDDDDDDDESRPIELPL-----ISSRQS 490
303 GVAVPHGDHYHFIYPSQMSLEERARIIPLRYRNMHWVPSRPSPQPTPEPSPGQ 362
491 LSEITPG-----PHSRKSTVSTRAADIP-----SDDEGYSDDSD-- 527
363 PAPNKKISNSGLSVQVLRKVGEGVPEKGISRYVFAKDLPSERVKNLEKSLSQSBSVS 422
528 -----DGNSS-----YTMENGEND--GDEDLNATYGYIQKLMLORSIS 566
423 HTLTAKENVAPRDOEFYDKAYNLTEAHKALFXNKGKNSDFQALDKLLERLNDESTNKE 482
567 IELASINELL--QDQCHDEYWNV-----NTSEIYSEFPDKNLRLTSQRE 611
483 KLVDDLLAFLAPITPERLQKPNQSI EYTEDEVRIAOADKYTTSQGYIFDEHDI---IS 539
612 KRM-----IAQMTKQDQDANNVWVWVQSVKDLMEMLV 640
540 DEGDAYVTFRMGHSHWIGKDSIDKEXVAAQAYTEKGLPPSPADYKAPPTGSAAI 599
641 DKDEKLVLA-----LDKERKLLKMLQK--LNNQPOVETEANEESDANS-- 683
600 YNRVAGEKRIPLVRLLPVVVEHTVEVKNGLIIPHKDHYNIKFAWFDHTYKAPNGYTL 659
684 -----MIKSGQESTN-----TLE 696
660 DLFAIKYVVEHDEPRHSNDGMASEHVLQK--DHSED--PNKNFADEEVPVETPAE 716
697 EIVKFIKATKESDD--SDADEFPAEBAAPSKKANDSEDLTTNKETPANAKKPCGEADED 754
717 PEV-----POVETEVKAQKE 733
755 ESLIVISSPOV--EKKNOQLKE 774

```

```

RESULT 49
UN89 CAEEL
ID UN89 CAEEL STANDARD; PRT; 6632 AA.
AC 001761; Q17362;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
GN UNC-89 OR C09D1.1.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditidae; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
RC STRAIN=Bristol N2;
RX MEDLINE=96180278; PubMed=8603916;
RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
assembly, encodes a giant modular protein composed of Ig and signal
transduction domains".
RT J. Cell Biol. 132:835-848(1996).
RL [2]
RN SEQUENCE FROM N.A.
RP STRAIN=Bristol N2;
RC Du Z., Le T.T., Wilson R.;
RA submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP REVISIONS.
RA Waterston R.;
RL submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Structural component of the muscle M-line. Myofibril
CC lattice assembly begins with positional cues laid down in the
CC basement membrane and muscle cell membrane. UNC-89 responds to
CC these signals, localizes, and then participates in assembling an
CC M-line.
CC -!- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
CC -!- SIMILARITY: Contains 1 DBP-homology (DB) domain.
CC -!- SIMILARITY: Contains 1 fibronectin type III domain.
CC -!- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 1 PH domain.
CC -!- SIMILARITY: Contains 5 RCD domain.
CC -!- SIMILARITY: Contains 1 SH3 domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC
CC -----
CC EMBL; U33058; AAB00542.1;
CC EMBL; AF003131; AAB54132.2;
CC PDB; 1FHO; 20-DEC-00.
CC WormPdb; C09D1.1; CE30426.
CC InterPro; IPR003961; FN III.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003598; Ig_C.
CC InterPro; IPR003006; Ig_MHC.
CC InterPro; IPR001849; PH.
CC InterPro; IPR007850; RCD.
CC InterPro; IPR000219; RhogEF.
CC InterPro; IPR001452; SH3.
CC Pfam; PF00041; fn3; 1.
CC Pfam; PF00047; Ig; 47.
CC Pfam; PF00169; PH; 1.
CC Pfam; PF05177; RCD; 5.
CC Pfam; PF00621; RhogEF; 1.
CC Pfam; PF00018; SH3; 1.
CC SMART; SM00408; IGC2; 23.
CC SMART; SM00325; RhogEF; 1.

```

FT	DISULFID	5508	5560	POTENTIAL.
FT	DISULFID	5616	5669	POTENTIAL.
FT	DISULFID	5722	5764	POTENTIAL.
FT	DISULFID	5836	5901	POTENTIAL.
FT	DISULFID	5946	5998	POTENTIAL.
FT	DISULFID	6036	6171	POTENTIAL.
FT	DISULFID	6421	6486	POTENTIAL.
FT	CONFLICT	2137	2137	A -> P (IN REF. 1).
FT	CONFLICT	2245	2247	AKA -> PKP (IN REF. 1).
FT	CONFLICT	2258	2258	E -> G (IN REF. 1).
FT	CONFLICT	2284	2284	M -> I (IN REF. 1).
FT	CONFLICT	2297	2297	A -> G (IN REF. 1).
FT	CONFLICT	3531	3531	DAGEY -> RRRI (IN REF. 1).
FT	CONFLICT	3884	3888	A -> V (IN REF. 1).
FT	CONFLICT	3929	3929	A -> P (IN REF. 1).
FT	CONFLICT	5134	5134	T -> S (IN REF. 1).
FT	CONFLICT	5145	5145	G -> A (IN REF. 1).
FT	CONFLICT	5185	5185	K -> N (IN REF. 1).
FT	CONFLICT	5199	5199	L -> F (IN REF. 1).
FT	CONFLICT	5202	5202	F -> L (IN REF. 1).
FT	CONFLICT	5213	5213	A -> G (IN REF. 1).
FT	CONFLICT	6178	6178	K -> B (IN REF. 1).
FT	CONFLICT	6268	6268	
SQ	SEQUENCE	6632	AA; 731665 MW; 262d3BED62960E89 CRC64;	
Query Match 3.2%; Score 134; DB 1; Length 6632;				
Best Local Similarity 19.8%; Pred. No. 1,1e+02;				
Matches 153; Conservative 113; Mismatches 239; Indels 208; Gaps 35;				
Oy	20 YIDGKATQKTENLPDEVSKREGINAEOIV--IKITDGGYTSHGDHYHYYNGKVVD	76		
Db	1078 YLKEELKQS-----DRVEIREPDGSIKISNIIKDIAGEIRA-----VATN	1120		
Oy	77 AIISEELMK-----DNYLKDEDIYNEVKGVIVYDGGKYUYLLDAAHADV--	126		
Db	1121 SEGSEDFKALIVQKKPFAPFDLRPSLVTE-KGSEAV-----FSMAHFGIPL	1168		
Oy	127 RTKE-EINROKOEHSOREEGTPRNDCAVALARSOGRYTTDDGYIFNAIDIETDAYI	185		
Db	1169 PTYMWSVNGKRKADQC-----EGA-----RVTDDESTVDGASILITDTAYYS	1211		
Oy	186 VPBGDHVHIYPKQELSASELAA-----AEAFSGRGSLNS-----RTYRRONS DN	231		
Db	1212 EVNHLTISTVAENTLGAEEFGAQTTEPKKESVIVEKODLSSEVGKEINAOVKASPEA	1271		
Oy	232 TSRNNWPPSYSNPGCTNTNTSNNST-----NSQSQSDNDISLKLQLYKLPLOSRAHV	285		
Db	1272 TTITIMETSILTSTTTTMTSTLEVTSTVGTVITFKESSESATVIIG-----GSGCVT	1325		
Oy	286 SDGLVPDPACITSTRAGVAVPRGDHNFIPYSOMSELERIARIIPLYRSNMHWPD SR	345		
Db	1326 EGISVSKEIEVSKITDSQTPVRG-----TPKRRVSAEBELPREVIDSRKKKKSBS PD	1380		
Oy	346 PEOPSPOTPEPBGPBPANLIKIDSNSLVSQLVRKVGEGVFEEEKGISRYVFAXDLPS	405		
Db	1381 KKEKSPKTEKRPASP-----TKKTGB-----EVMS-----PK	1408		
Oy	406 ETYNGLSKISKQESVHTLTAKENVAPRDQEFYDKAVLLTBANKALFXNKGRNSDFQ	465		
Db	1409 EKSPASPTKKEKSPAABEVKSPTKKEKSPSPPTKKEKSPSPPTK-----K	1453		
Oy	466 ALDLRLRLNDES-TNNEKXLDVLLFLAITHPERLGRKNSQIEYTEDEVRILAQLADX	524		
Db	1454 TGDBVEKESPPKSTTKKEKSP-----KPEDVXSVPVK-EKSPDXTNVIVSSSET	1502		
Oy	525 TTSDGYIFDEHDITSDEGDAVVTPHMGHSHMICKDSLDEKXYAAQAQVYTKEGKLPSPSD	584		
Db	1503 TI-----EXTETMTTEMTHBEEBSERTSVK-KEKTIPEKVEDK-----PSSPT	1543		
Oy	585 ADVKANPTGSAAIIVNRVGEKRIPLVRLPYWEH-----TYEVKNGNL-IIPHGDHYH	638		
Db	1544 KKOK-SPEKSIITEBIKSPVKKES-----PERVEEKEPASPTKKEKSPKPPASPTKSEN	1596		

```

OY 639 NIKFAMPDHTYKANGTYLDELFAITIKYVEHPDERPHSDNGWGNASEHVLGKKDHSB 698
DB 1597 EKXSPJTKKE---KXSEKSVVELKSPKESPKADKPKSPKTKKSPK-----KSATED 1648
OY 699 ---PKNKRKADEEPEVEETPAEPEVPOVETKEVLAOLKEAEVLLAAYTSSSLKA 748
DB 1649 VKSPYKKEKSPK-VEEKPTSP-----TKKESSEPTK-----TDDEVKS 1686

RESULT 50
LMGI_MOUSE STANDARD; PRT: 1607 AA.
ID LMGI_MOUSE
AC 21-JUL-1986 (Rel. 01, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Laminin gamma-1 chain precursor (Laminin B2 chain).
GN LAMC1 OR LAMC-1 OR LAMB-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=88059118; PubMed=3680290;
RA Sasaki M., Yamada Y.;
RT "Type laminin B2 chain has a multidomain structure homologous to the
RT B1 chain."
RL J. Biol. Chem. 262:17111-17117(1987).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=89000737; PubMed=3167041;
RA Durkin M.E., Barros B.B., Liu S.-H., Phillips S.L., Chung A.E.;
RT "Primary structure of the mouse laminin B2 chain and comparison with
RT laminin B1."
RL Biochemistry 27:5198-5204(1988).
RN (3)
RP SEQUENCE OF 1-239 FROM N.A.
RX MEDLINE=88228071; PubMed=2836421;
RA Ogawa K., Burbelo P.D., Sasaki M., Yamada Y.;
RT "The laminin B2 chain promoter contains unique repeat sequences and
RT is active in transient transfection."
RL J. Biol. Chem. 263:8384-8389(1988).
RN (4)
RP SEQUENCE OF 1391-1607 FROM N.A.
RX MEDLINE=85051302; PubMed=6209134;
RA Bartlow D.P., Green N.M., Kurkinen M., Hogan B.L.M.;
RT "Sequencing of laminin B chain cDNAs reveals C-terminal regions of
RT coiled-coil alpha-helix."
RL EMBO J. 3:2355-2362(1984).
RN (5)
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 771-932.
RX MEDLINE=96196434; PubMed=8648630;
RA Stetefeld J., Mayer U., Timpl R., Huber R.;
RT "Crystal structure of three consecutive laminin-type epidermal growth
RT factor-like (LE) modules of laminin gamma1 chain harboring the
RT nitrogen binding site."
RL J. Mol. Biol. 257:644-657(1996).
RN (6)
RP STRUCTURE BY NMR OF 824-881.
RX MEDLINE=96196435; PubMed=8648631;
RA Baumgartner R., Czisch M., Mayer U., Poeschl E., Huber R.,
RA Timpl R., Holak T.A.;
RT "Structure of the nitrogen binding LE module of the laminin gamma1
RT chain in solution."
RL J. Mol. Biol. 257:658-668(1996).
CC -1- FUNCTION: Binding to cells via a high affinity receptor, laminin
CC is thought to mediate the attachment, migration, and organization
CC of cells into tissues during embryonic development by interacting
CC with other extracellular matrix components.
CC -1- SUBUNIT: Laminin is a complex glycoprotein, consisting of three
CC different polypeptide chains (alpha, beta, gamma), which are bound
CC to each other by disulfide bonds into a cross-shaped molecule

```

```

CC CC Comprising one long and three short arms with globules at each
CC end.
CC THE GAMMA-1 CHAIN IS A SUBUNIT OF LAMININ-1 (EHS LAMININ),
CC LAMININ-2 (MEROSIN), LAMININ-3 (S-LAMININ), LAMININ-4 (S-MEROSIN),
CC LAMININ-6 (K-LAMININ) AND LAMININ-7 (KS-LAMININ).
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC -1- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES (MAJOR
CC COMPONENT).
CC -1- DOMAIN: THE ALPHA-HELICAL DOMAINS I AND II ARE THOUGHT TO INTERACT
CC WITH OTHER LAMININ CHAINS TO FORM A COILED COIL STRUCTURE.
CC -1- DOMAIN: DOMAINS VI AND IV ARE GLOBULAR.
CC -1- SIMILARITY: Contains 1 laminin N-terminal domain.
CC -1- SIMILARITY: Contains 1 laminin EGF-like domains.
CC -1- SIMILARITY: Contains 1 laminin IV domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL, X05211; CAA28838.1; -
CC DR EMBL, J03484; AAA39405.1; -
CC DR EMBL, J02930; AAA39408.1; -
CC DR EMBL, J03749; AAA39409.1; -
CC DR PIR, A28469; MMSB2.
CC DR PDB, 1KLO; 20-AUG-97.
CC DR PDB, 1TLE; 12-FEB-97.
CC DR MGD, MGI:99914; Lamc1.
CC DR GO, GO:0005604; C:basement membrane; IDA.
CC DR InterPro: IPR006209; EGF like.
CC DR InterPro: IPR000034; Laminin B.
CC DR InterPro: IPR002049; Laminin_EGF.
CC DR InterPro: IPR001886; LaminT.
CC DR Pfam, PF00052; laminin_B; 1.
CC DR Pfam, PF00053; laminin_EGF; 10.
CC DR Pfam, PF00055; laminin_Nterm; 1.
CC DR PRINTS, PR00011; EGF_LAMININ.
CC DR ProDom, PD002082; Lam_N2; 1.
CC DR SMART, SM00180; EGF_Lam; 8.
CC DR SMART, SM00281; Lamb; 1.
CC DR SMART, SM00136; LaminT; 1.
CC DR PROSITE, PS00022; EGF_1; 8.
CC DR PROSITE, PS01248; LAMININ_Type_EGF; 10.
CC DR Glycoprotein; Basement membrane; Extracellular matrix; Coiled coil;
CC KW Laminin EGF-like domain; Cell adhesion; Repeat; Signal; 3D-structure.
CC FT SIGNAL 1 33
CC FT CHAIN 34 1607 LAMININ GAMMA-1 CHAIN.
CC FT DOMAIN 34 283 LAMININ N-TERMINAL (DOMAIN VI).
CC FT DOMAIN 284 339 LAMININ EGF-LIKE 1.
CC FT DOMAIN 340 395 LAMININ EGF-LIKE 2.
CC FT DOMAIN 396 442 LAMININ EGF-LIKE 3.
CC FT DOMAIN 443 492 LAMININ EGF-LIKE 4.
CC FT DOMAIN 493 502 LAMININ EGF-LIKE 5 (N-TERMINAL).
CC FT DOMAIN 503 687 LAMININ DOMAIN IV.
CC FT DOMAIN 688 721 LAMININ EGF-LIKE 5 (C-TERMINAL).
CC FT DOMAIN 722 770 LAMININ EGF-LIKE 6.
CC FT DOMAIN 771 825 LAMININ EGF-LIKE 7.
CC FT DOMAIN 826 881 LAMININ EGF-LIKE 8 (NIDOGEN-BINDING).
CC FT DOMAIN 882 932 LAMININ EGF-LIKE 9.
CC FT DOMAIN 933 980 LAMININ EGF-LIKE 10.
CC FT DOMAIN 981 1028 LAMININ EGF-LIKE 11.
CC FT DOMAIN 1029 1607 DOMAIN II AND I.
CC FT DOMAIN 1030 1594 COILED COIL (POTENTIAL).
CC FT DISULFD 340 349 BY SIMILARITY.
CC FT DISULFD 342 365 BY SIMILARITY.
CC FT DISULFD 368 377 BY SIMILARITY.
CC FT DISULFD 380 393 BY SIMILARITY.
CC FT DISULFD 396 408 BY SIMILARITY.
CC FT DISULFD 398 414 BY SIMILARITY.

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 09:54:32 ; Search time 89 Seconds
(without alignments)
2307.976 Million cell updates/sec

Title: US-09-765-271-56
Perfect score: 4165
Sequence: 1 SYELGLYQARTVKNRVS.....KLALIKGSPSSVSKEKIN 796

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database :
1: SP archaea:*
2: SP bacteria:*
3: SP fungi:*
4: SP human:*
5: SP invertebrate:*
6: SP mammal:*
7: SP mhc:*
8: SP organelle:*
9: SP phage:*
10: SP plant:*
11: SP rodent:*
12: SP virus:*
13: SP vertebrate:*
14: SP unclassified:*
15: SP virus:*
16: SP bacteriap:*
17: SP archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4163	100.0	802	16 Q9TQW6	Q9TQW6 streptococc
2	4163	100.0	816	2 Q9AHT9	Q9AHT9 streptococc
3	4159	99.9	828	16 Q8BPQ2	Q8BPQ2 streptococc
4	3226	77.7	844	2 Q9AG74	Q9AG74 streptococc
5	3224	77.4	855	16 Q8CWR4	Q8CWR4 streptococc
6	2797.5	67.2	819	16 Q9TQW6	Q9TQW6 streptococc
7	2786.5	66.9	819	2 Q9ANY3	Q9ANY3 streptococc
8	2772	66.6	839	16 Q9ANY2	Q9ANY2 streptococc
9	2733	65.6	853	16 Q8DQ08	Q8DQ08 streptococc
10	1246	29.9	1039	16 Q9ANY1	Q9ANY1 streptococc
11	1243	29.8	1039	16 Q8DQ07	Q8DQ07 streptococc
12	949	22.8	825	16 Q9XV4	Q9XV4 streptococc
13	946	22.7	823	16 Q8R282	Q8R282 streptococc
14	945	22.7	822	16 Q8F4U1	Q8F4U1 streptococc
15	944	22.7	825	2 Q9JG75	Q9JG75 streptococc
16	940	22.6	823	16 Q8K5Q1	Q8K5Q1 streptococc

17	937	22.5	822	16 Q8D281	Q8D281 streptococc
18	934	22.4	822	2 Q9ZHG7	Q9ZHG7 streptococc
19	916.5	22.0	481	16 Q8E338	Q8E338 streptococc
20	679.5	16.3	289	2 Q9AE21	Q9AE21 streptococc
21	305	7.3	877	16 Q8E5R2	Q8E5R2 streptococc
22	302	7.3	877	16 Q8E029	Q8E029 streptococc
23	243	5.8	182	16 Q8DQ06	Q8DQ06 streptococc
24	223	5.4	792	16 Q9Z276	Q9Z276 streptococc
25	213	5.1	792	16 Q8K714	Q8K714 streptococc
26	197.5	4.7	1078	5 Q963T1	Q963T1 streptococc
27	192.5	4.6	1390	5 Q77033	Q77033 dictyosteli
28	192	4.6	792	16 Q8P0G5	Q8P0G5 streptococc
29	181.5	4.4	2752	5 Q8ICT8	Q8ICT8 streptococc
30	178	4.3	565	5 Q15754	Q15754 dictyosteli
31	178	4.3	1233	5 Q8J556	Q8J556 dictyosteli
32	177	4.2	8591	5 Q8IB94	Q8IB94 plasmodium
33	176.5	4.2	1271	5 Q25860	Q25860 plasmodium
34	175.5	4.2	1236	5 Q9GTX2	Q9GTX2 plasmodium
35	175.5	4.2	5507	5 Q8IHJ3	Q8IHJ3 plasmodium
36	173	4.2	1063	5 Q8ICV5	Q8ICV5 plasmodium
37	172.5	4.1	5458	5 Q9U459	Q9U459 plasmodium
38	172	4.1	1795	16 Q9ICJ9	Q9ICJ9 streptococc
39	172	4.1	2478	2 Q9ICJ2	Q9ICJ2 streptococc
40	172	4.1	2481	16 Q9QOR6	Q9QOR6 streptococc
41	171.5	4.1	3452	5 Q8IEA3	Q8IEA3 plasmodium
42	170	4.1	2478	2 Q9RL69	Q9RL69 streptococc
43	168.5	4.0	1043	10 Q82345	Q82345 arabisdopsis
44	168.5	4.0	1708	5 Q8I413	Q8I413 plasmodium
45	168.5	4.0	2651	5 Q8IEB6	Q8IEB6 plasmodium
46	166	4.0	1465	5 Q8I223	Q8I223 plasmodium
47	166	4.0	1826	5 Q97255	Q97255 plasmodium
48	166	4.0	16215	5 Q9NFS3	Q9NFS3 drosophila
49	166	4.0	18074	5 Q9I7U4	Q9I7U4 drosophila
50	165.5	4.0	1373	5 Q8I3A8	Q8I3A8 plasmodium
51	165.5	4.0	2227	5 Q8I121	Q8I121 plasmodium
52	165.5	4.0	3381	5 Q8I2V4	Q8I2V4 plasmodium
53	165	4.0	891	16 Q99TD3	Q99TD3 streptococc
54	165	4.0	3370	5 Q8I1N9	Q8I1N9 plasmodium
55	164	3.9	891	16 Q931P4	Q931P4 streptococc
56	162.5	3.9	1488	5 Q8IDP7	Q8IDP7 plasmodium
57	162.5	3.9	2647	5 Q9U4X0	Q9U4X0 plasmodium
58	162	3.9	895	16 Q8NW39	Q8NW39 streptococc
59	162	3.9	2467	5 Q8I1D3	Q8I1D3 plasmodium
60	162	3.9	2910	10 Q9FND5	Q9FND5 arabisdopsis
61	162	3.9	3504	5 Q8IL45	Q8IL45 plasmodium
62	162	3.9	4405	5 Q8IL22	Q8IL22 plasmodium
63	162	3.9	5322	5 Q9VPL9	Q9VPL9 drosophila
64	161.5	3.9	1003	5 Q8I659	Q8I659 plasmodium
65	160.5	3.9	1083	5 Q8IBL1	Q8IBL1 plasmodium
66	160.5	3.9	1461	5 Q8I5H0	Q8I5H0 plasmodium
67	160	3.8	3692	16 Q8CNU9	Q8CNU9 streptococc
68	159.5	3.8	1129	16 Q8XLI1	Q8XLI1 streptococc
69	159	3.8	9439	16 Q8CP76	Q8CP76 streptococc
70	158	3.8	1650	5 Q77328	Q77328 plasmodium
71	158	3.8	1840	5 Q8IED3	Q8IED3 plasmodium
72	157	3.8	1859	5 Q8IC27	Q8IC27 plasmodium
73	156.5	3.8	1139	5 Q8IJS7	Q8IJS7 plasmodium
74	156.5	3.8	1157	10 Q8H6X1	Q8H6X1 arabisdopsis
75	156.5	3.8	1185	5 Q8IMF5	Q8IMF5 drosophila
76	156.5	3.8	1495	10 Q9I2A8	Q9I2A8 arabisdopsis
77	156	3.7	519	10 Q9SIG8	Q9SIG8 arabisdopsis
78	156	3.7	1692	5 Q8IJD6	Q8IJD6 plasmodium
79	155.5	3.7	1038	3 Q8PB19	Q8PB19 schizosacch
80	155.5	3.7	1298	5 Q9VEU8	Q9VEU8 drosophila
81	155.5	3.7	1413	11 Q8CG83	Q8CG83 mus musculu
82	155.5	3.7	3069	5 Q8I350	Q8I350 plasmodium
83	155	3.7	869	4 Q9NYF8	Q9NYF8 homo sapien
84	155	3.7	920	4 Q14673	Q14673 homo sapien
85	155	3.7	1779	5 Q8I1W2	Q8I1W2 plasmodium
86	154.5	3.7	1338	5 Q77306	Q77306 plasmodium
87	154.5	3.7	1463	2 Q8E919	Q8E919 streptococc
88	154.5	3.7	1516	5 Q96154	Q96154 plasmodium
89	154	3.7	749	5 Q02602	Q02602 plasmodium

90 154 3.7 4723 5 O8MU8 Q8MU8 drosophila
 91 153.5 3.7 1019 5 O8IDP8 O8IDP8 plasmidium
 92 153.5 3.7 1881 16 O8RGK2 O8RGK2 fuscobacteri
 93 153.5 3.7 1898 5 O8ILA3 O8ILA3 plasmidium
 94 153.5 3.7 2867 5 O9N2M3 O9N2M3 plasmidium
 95 153 3.7 1558 5 O8ILU5 O8ILU5 plasmidium
 96 153 3.7 2402 13 O9DEH4 O9DEH4 gallus galli
 97 153 3.7 2413 5 O8ICH7 O8ICH7 plasmidium
 98 152.5 3.7 1434 5 O8IUI3 O8IUI3 plasmidium
 99 152.5 3.7 1928 5 O8IUI1 O8IUI1 plasmidium
 100 152.5 3.7 5251 5 O8IUD4 O8IUD4 plasmidium

ALIGNMENTS

RESULT 1

O97OM8 PRELIMINARY; PRT; 802 AA.
 ID O97OM8
 AC O97OM8:
 DT 01-OCT-2001 (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Conserved domain protein.
 GN SP1175.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 CC Streptococcus.
 OX NCBI_TaxID=1313;
 RN SEQUENCE FROM N.A.
 RP STRAIN=TIGR4;
 RC MEDLINE=21357209; PubMed=11463916;
 RX Tetteelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
 Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.J.,
 Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
 Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
 Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
 McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
 Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
 RA Dougherty B.A., Mortison D.A., Hollingshead S.K., Frazer C.M.;
 RA "Complete genome sequence of a virulent isolate of Streptococcus
 RT pneumoniae.";
 RL Science 293:498-506(2001).
 DR EMBL: AE007418; AK75284.1; -.
 DR TIGR: SP1175; -.
 DR InterPro: IPR006270; Strep_his_triad.
 DR TIGRFAMS: TIGR01363; strep_his_triad; 2.
 KW Complete proteome.
 SQ SEQUENCE 802 AA; 90080 MW; 45ECB8364EFA1833 CRC64;

Query Match 100.0%; Score 4163; DB 16; Length 802;
 Best Local Similarity 99.9%; Pred. No. 1.6e-229;
 Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 SYELGLVQARTYKNNRVSYIDGKATOKTENTLTPDEVSKREGINAEOIVIKITDOGYVT 60
 DB 7 SYELGLVQARTYKNNRVSYIDGKATOKTENTLTPDEVSKREGINAEOIVIKITDOGYVT 66
 OY 61 SHGDHYHYNGKVPDAIIESELLMKDPYKLDKEDIVNENGVGVYIKVDGKYVYLLDA 120
 DB 67 SHGDHYHYNGKVPDAIIESELLMKDPYKLDKEDIVNENGVGVYIKVDGKYVYLLDA 126
 OY 121 AHADVNRTEINRQKOEHSQRECGTPRNDGAVALAASQGRYTTDDGIYFNASDIIDT 180
 DB 127 AHADVNRTEINRQKOEHSQRECGTPRNDGAVALAASQGRYTTDDGIYFNASDIIDT 186
 OY 181 GDAYVVPBGDHYHYPKNELSASELAAAEFLSGGINSNSTYROROSDNTRNNWPS 240
 DB 187 GDAYVVPBGDHYHYPKNELSASELAAAEFLSGGINSNSTYROROSDNTRNNWPS 246
 OY 241 VSNPGTTNTNTSNNSTNSQASQNDIDSLKQLYKLPISQRHVESDGLVDPDAITSTRT 300

DB 247 VSNPGTTNTNTSNNSTNSQASQNDIDSLKQLYKLPISQRHVESDGLVDPDAITSTRT 306
 OY 301 ARGVAVPHGDHYHYPIYQSMSELEERTARIIPLYRSHNHWPDSRPPSPQPPSPSPG 360
 DB 307 ARGVAVPHGDHYHYPIYQSMSELEERTARIIPLYRSHNHWPDSRPPSPQPPSPSPG 366
 OY 361 POPAPNKLIDNSNLVLQVRKVGEGYVFEKGISRVVPAKDLPSFTVNKLSKLSKQES 420
 DB 367 POPAPNKLIDNSNLVLQVRKVGEGYVFEKGISRVVPAKDLPSFTVNKLSKLSKQES 426
 OY 421 VSHTLTKKENVAPRDEPFYDKAYNLLTEAHKALPKXKGRNSDFQALDKLERLNDSTN 480
 DB 427 VSHTLTKKENVAPRDEPFYDKAYNLLTEAHKALPKXKGRNSDFQALDKLERLNDSTN 486
 OY 481 KEKLVDDLAFIAPITPERLGRKNSQIETDEVRFAQLADKTTSDGYTFDEHDIISD 540
 DB 487 KEKLVDDLAFIAPITPERLGRKNSQIETDEVRFAQLADKTTSDGYTFDEHDIISD 546
 OY 541 EGDAYVTPHMGSHWICKDSLSDKEKYAAQAYTEKGIPLPSPPADYKANPTGDSAAAIY 600
 DB 547 EGDAYVTPHMGSHWICKDSLSDKEKYAAQAYTEKGIPLPSPPADYKANPTGDSAAAIY 606
 OY 601 NRKVGKRIPLVRLPYVVEHTVEYKGNLII PKRDHYHNIFKAFWDDHTYKAPNGYTLSD 660
 DB 607 NRKVGKRIPLVRLPYVVEHTVEYKGNLII PKRDHYHNIFKAFWDDHTYKAPNGYTLSD 666
 OY 661 LFATIKYVVEHPDRPHSNDGMSNASEHVIGKDHSDPNKFNPADEPVEETPAEPVP 720
 DB 667 LFATIKYVVEHPDRPHSNDGMSNASEHVIGKDHSDPNKFNPADEPVEETPAEPVP 726
 OY 721 QVETEKYEAQLKEAEVLLAKVTSSLRKANATETLAGLRNNLTLOIMDNNSIMAEAKLLA 780
 DB 727 QVETEKYEAQLKEAEVLLAKVTSSLRKANATETLAGLRNNLTLOIMDNNSIMAEAKLLA 786
 OY 781 LKGSNPSVSSEKXIN 796
 DB 787 LKGSNPSVSSEKXIN 802

RESULT 2

O9AHT9 PRELIMINARY; PRT; 816 AA.
 ID O9AHT9
 AC O9AHT9:
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Pneumococcal histidine triad A protein.
 GN PHTA.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 CC Streptococcus.
 OX NCBI_TaxID=1313;
 RN SEQUENCE FROM N.A.
 RP STRAIN=N4;
 RC MEDLINE=21116976; PubMed=11179332;
 RX Wiemann T.M., Heinrichs J.H., Adamou J.E., Erwin A.L., Kunsch C.,
 Choi G.H., Barash S.C., Rosen C.A., Maure H.R., Tuomanen E.,
 RA Gayle A., Brawsh Y.A., Walsh W., Barron P., Lathigra R., Hanson M.,
 RA Langemann S., Johnson S., Koenig S.;
 RA "Use of a Whole Genome Approach To Identify Vaccine Molecules
 RT Affording Protection Against Streptococcus pneumoniae Infection.";
 RL Infect. Immun. 69:1593-1598(2001).
 DR EMBL: AF291695; AK19155.1; -.
 DR InterPro: IPR006270; Strep_his_triad.
 DR TIGRFAMS: TIGR01363; strep_his_triad; 2.
 SQ SEQUENCE 816 AA; 91519 MW; 5359126A611D27ED CRC64;

Query Match 100.0%; Score 4163; DB 2; Length 816;
 Best Local Similarity 100.0%; Pred. No. 1.6e-229;
 Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 SYELGLVQARTYKNNRVSYIDGKATOKTENTLTPDEVSKREGINAEOIVIKITDOGYVT 60

```

Db      21 SYELGLYQARTVKENNRVSYIDGKATQKTEMLTPDEVSKREGINAEQIVIKITDGGYVT 80
QY      61 SHGDHYHYNGKVPYDAIISEELMKDPNYKLKDEDIVNEVKGYYIKVDGKYVYLKDA 120
Db      81 SHGDHYHYNGKVPYDAIISEELMKDPNYKLKDEDIVNEVKGYYIKVDGKYVYLKDA 140
QY      121 AAADVNRTEKEINROKQESHQREGGTPRNDGVALARSGRYTTDDGYIFNADIIEDT 180
Db      141 AAADVNRTEKEINROKQESHQREGGTPRNDGVALARSGRYTTDDGYIFNADIIEDT 200
QY      181 GDAVIVPHGDHYHYIPKNELASSELAAAEFLSGRGLNSRTRYRQNSDNTSRITWVPS 240
Db      201 GDAVIVPHGDHYHYIPKNELASSELAAAEFLSGRGLNSRTRYRQNSDNTSRITWVPS 260
QY      241 VSNPQTTNTNTSNSTNSQASQNSNDISLKLQYLKPLSQRHVESDGLVFPDQAQITST 300
Db      261 VSNPQTTNTNTSNSTNSQASQNSNDISLKLQYLKPLSQRHVESDGLVFPDQAQITST 320
QY      301 ARGVAAPHGDHYHYIPYQSMSELEERARIIPLRYSNMHWVPSRPEQSPQPTPEPSPG 360
Db      321 ARGVAAPHGDHYHYIPYQSMSELEERARIIPLRYSNMHWVPSRPEQSPQPTPEPSPG 380
QY      361 POPAPNLKIDSNSLSVQLVRKYGEGYVEEKGISRYVFAKDLPSSTVKNLESKLSKQES 420
Db      381 POPAPNLKIDSNSLSVQLVRKYGEGYVEEKGISRYVFAKDLPSSTVKNLESKLSKQES 440
QY      421 VSHITLTAKKENVAPRQDEFYDKAYNLTTEAHKALFYNKGRNSDFQALDKLERLNDSTN 480
Db      441 VSHITLTAKKENVAPRQDEFYDKAYNLTTEAHKALFYNKGRNSDFQALDKLERLNDSTN 500
QY      481 KEKLVDDLAFAPLITHPERLQKPNQIETDEVERIAQADKYTTSDGYIFPEHDIISD 540
Db      501 KEKLVDDLAFAPLITHPERLQKPNQIETDEVERIAQADKYTTSDGYIFPEHDIISD 560
QY      541 EGDAYVTPHMGSHHWIGKDSLSDKEKVAQAQATYKEKILPPSPDADVKANPTGDSAAAY 600
Db      561 EGDAYVTPHMGSHHWIGKDSLSDKEKVAQAQATYKEKILPPSPDADVKANPTGDSAAAY 620
QY      601 NRYKGEKRIPLVRLPYMVEHTVEVKNGNLIIPHKDHYHNIKFPAMFDHTYKAPNGYTLED 660
Db      621 NRYKGEKRIPLVRLPYMVEHTVEVKNGNLIIPHKDHYHNIKFPAMFDHTYKAPNGYTLED 680
QY      661 LPAITIKYYEHPERPHSNDGNGNASEHVLGKKDHSDPNKFKADEEVEETPAPEVP 720
Db      681 LPAITIKYYEHPERPHSNDGNGNASEHVLGKKDHSDPNKFKADEEVEETPAPEVP 740
QY      721 QVETKEVAQLKEAEVLLAKVTDSILKANATETLAQRNNLTQIMDNNSIMAEAEKLLA 780
Db      741 QVETKEVAQLKEAEVLLAKVTDSILKANATETLAQRNNLTQIMDNNSIMAEAEKLLA 800
QY      781 LKGSNPPSSVSKEKIN 796
Db      801 LKGSNPPSSVSKEKIN 816

RESULT 3
Q8DPQ2 ID Q8DPQ2 PRELIMINARY; PRT; 828 AA.
AC Q8DPQ2;
DT 01-MAR-2003 (TEMBLrel. 23, Created)
DT 01-MAR-2003 (TEMBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
DE Pneumococcal histidine triad protease A.
GN PHTA OR SPRI061.
OC Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
Hoskins J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burett S.,

```

```

RA Dehoff B.S., Estrem S.T., Fritz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA Leblanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAben S.M., McHenry M., Mcleaster K., Mundy C.W., Nicot T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rokey P.,
RA Sun P.-W., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostek P.R. Jr., Skatrud P.L.,
RA Glass J.I.,
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.",
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008479; AAK99865.1; -.
KW Complete proteome.
SQ SEQUENCE 828 AA; 93015 MW; 12CCCF407B550C1D CRC64;

Query Match          99.9%; Score 4159; DB 16; Length 828;
Best Local Similarity 99.7%; Pred. No. 2,8e+229;
Matches 794; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      1 SYELGLYQARTVKENNRVSYIDGKATQKTEMLTPDEVSKREGINAEQIVIKITDGGYVT 60
Db      33 SYELGLYQARTVKENNRVSYIDGKATQKTEMLTPDEVSKREGINAEQIVIKITDGGYVT 92
QY      61 SHGDHYHYNGKVPYDAIISEELMKDPNYKLKDEDIVNEVKGYYIKVDGKYVYLKDA 120
Db      93 SHGDHYHYNGKVPYDAIISEELMKDPNYKLKDEDIVNEVKGYYIKVDGKYVYLKDA 152
QY      121 AAADVNRTEKEINROKQESHQREGGTPRNDGVALARSGRYTTDDGYIFNADIIEDT 180
Db      153 AAADVNRTEKEINROKQESHQREGGTPRNDGVALARSGRYTTDDGYIFNADIIEDT 212
QY      181 GDAVIVPHGDHYHYIPKNELASSELAAAEFLSGRGLNSRTRYRQNSDNTSRITWVPS 240
Db      213 GDAVIVPHGDHYHYIPKNELASSELAAAEFLSGRGLNSRTRYRQNSDNTSRITWVPS 272
QY      241 VSNPQTTNTNTSNSTNSQASQNSNDISLKLQYLKPLSQRHVESDGLVFPDQAQITST 300
Db      273 VSNPQTTNTNTSNSTNSQASQNSNDISLKLQYLKPLSQRHVESDGLVFPDQAQITST 332
QY      301 ARGVAAPHGDHYHYIPYQSMSELEERARIIPLRYSNMHWVPSRPEQSPQPTPEPSPG 360
Db      333 ARGVAAPHGDHYHYIPYQSMSELEERARIIPLRYSNMHWVPSRPEQSPQPTPEPSPG 392
QY      361 POPAPNLKIDSNSLSVQLVRKYGEGYVEEKGISRYVFAKDLPSSTVKNLESKLSKQES 420
Db      393 POPAPNLKIDSNSLSVQLVRKYGEGYVEEKGISRYVFAKDLPSSTVKNLESKLSKQES 452
QY      421 VSHITLTAKKENVAPRQDEFYDKAYNLTTEAHKALFYNKGRNSDFQALDKLERLNDSTN 480
Db      453 VSHITLTAKKENVAPRQDEFYDKAYNLTTEAHKALFYNKGRNSDFQALDKLERLNDSTN 512
QY      481 KEKLVDDLAFAPLITHPERLQKPNQIETDEVERIAQADKYTTSDGYIFPEHDIISD 540
Db      513 KEKLVDDLAFAPLITHPERLQKPNQIETDEVERIAQADKYTTSDGYIFPEHDIISD 572
QY      541 EGDAYVTPHMGSHHWIGKDSLSDKEKVAQAQATYKEKILPPSPDADVKANPTGDSAAAY 600
Db      573 EGDAYVTPHMGSHHWIGKDSLSDKEKVAQAQATYKEKILPPSPDADVKANPTGDSAAAY 632
QY      601 NRYKGEKRIPLVRLPYMVEHTVEVKNGNLIIPHKDHYHNIKFPAMFDHTYKAPNGYTLED 660
Db      633 NRYKGEKRIPLVRLPYMVEHTVEVKNGNLIIPHKDHYHNIKFPAMFDHTYKAPNGYTLED 692
QY      661 LPAITIKYYEHPERPHSNDGNGNASEHVLGKKDHSDPNKFKADEEVEETPAPEVP 720
Db      693 LPAITIKYYEHPERPHSNDGNGNASEHVLGKKDHSDPNKFKADEEVEETPAPEVP 752
QY      721 QVETKEVAQLKEAEVLLAKVTDSILKANATETLAQRNNLTQIMDNNSIMAEAEKLLA 780
Db      753 QVETKEVAQLKEAEVLLAKVTDSILKANATETLAQRNNLTQIMDNNSIMAEAEKLLA 812
QY      781 LKGSNPPSSVSKEKIN 796
Db      813 LKGSNPPSSVSKEKIN 828

```

RESULT 4

Q9AG74 PRELIMINARY; PRT; 844 AA.
 ID Q9AG74
 AC Q9AG74;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE PhpA.
 GN PhpA.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OC NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21246685; PubMed=11349048;
 RA Zhang Y., Mast A.W., Barniak V., Mountzouras K., Hostetter M.K.,
 RA Green B.A.;
 RT "Recombinant PhpA protein, a Unique Histidine Motif-Containing Protein
 RT from Streptococcus pneumoniae, Protects Mice against Intranasal
 RT Pneumococcal Challenge."
 RL Infect. Immun. 69:3827-3836(2001).
 DR EMBL; AF340221; AAK26629.1; -
 DR InterPro: IPR006270: Strept_his_triad.
 DR TrEMBL; TIGR01363: strep_his_triad.2
 SQ SEQUENCE 844 AA; 94769 MW; D738A55290F89902 CRC64;

Query Match 77.7%; Score 3236; DB 2; Length 844;
 Best Local Similarity 75.3%; Pred. No. 1.5e-176;
 Matches 620; Conservative 66; Mismatches 99; Indels 38; Gaps 4;

QY 1 SYELGLYQARTYKNNRVSITDGKATQKTEMLTPDEVSKREGINAEQIVIKITDQGYVT 60
 DB 21 SYELGLYQARTYKNNRVSITDGKATQKTEMLTPDEVSKREGINAEQIVIKITDQGYVT 80
 QY 61 SHGHHYHYNGKVPYDAIIEBELMKDPNYKLKDEDIVNEVGVYIVKDGKYVYLKDA 120
 DB 81 SHGHHYHYNGKVPYDAIIEBELMKDPNYKLKDEDIVNEVGVYIVKDGKYVYLKDA 140
 QY 121 AHAADVRTKEELNRKQEHSHQREGGTPRNDGAVALARASQGRYTTDDGYIFNADIIEDT 180
 DB 141 AHAADVRTKEELNRKQEHSHQREGGTPRNDGAVALARASQGRYTTDDGYIFNADIIEDT 200
 QY 181 GDAYIVPHGDHYHIFPKNELSASELAALAEFLSGGNLSNRTYTRQNSDNTSRNWPVS 240
 DB 201 GDAYIVPHGDHYHIFPKNELSASELAALAEFLSGGNLSNRTYTRQNSDNTSRNWPVS 260
 QY 241 VSNPCTNTNTSNNSNTNSQASQSDNDISLKLQYLKPLSQRHVSDDGIVPDAQITSRT 300
 DB 261 VSNPCTNTNTSNNSNTNSQASQSDNDISLKLQYLKPLSQRHVSDDGIVPDAQITSRT 320
 QY 301 ARGVAVPHGDHYHIFPKNELSASELAALAEFLSGGNLSNRTYTRQNSDNTSRNWPVS 360
 DB 321 ARGVAVPHGDHYHIFPKNELSASELAALAEFLSGGNLSNRTYTRQNSDNTSRNWPVS 380
 QY 361 POPAPNKL-IDSN--SSLVSQLVRKVGEGYFEEKGISRYFAADLPSEYVKNLESTLS 416
 DB 381 POPAPNPOPAPSNPIDEKLVKVAARKVDGYVEENGVPARYIPADDLAETAGIDSKLA 440
 QY 417 KOESVSHLTAKKEVVARPDQEFYKAVNLLTEAKALFXNNGRNSDQALDKLERLND 476
 DB 441 KOESVSHLTAKKEVVARPDQEFYKAVNLLTEAKALFXNNGRNSDQALDKLERLND 500
 QY 477 ESTNKEKLVDLLAFAPITPERLQKNSQIEYEDVRIQLADKXTTSDGYTFDEHD 536
 DB 501 VSSDKVKLVDDILARLIRHERIGKXNAQITTYDDDELQYAKLAGKXTTBDGYTFDDR 560
 QY 537 IISDEGDAYVTPHMGSHWIGKDSLDEKVAQAQYTYEKGILPPSPADYKANFTGSA 596
 DB 561 ITSDEGDAYVTPHMGSHWIKDKSLSEAEARAAQAQYAKKEGILTPPSTDHODSGNTEAKGA 620

RESULT 5

Q8CWR4 PRELIMINARY; PRT; 855 AA.
 ID Q8CWR4
 AC Q8CWR4;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Histidine motif-containing protein.
 GN PHPA OR SPR1060.
 OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OC NCBI_TaxID=171101;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21429245; PubMed=11544234;
 RA Hoskins J., Albom W.E., Jr., Arnold J., Blaszczak L.C., Burgette S.,
 RA Dehoff B.S., Estrem S.T., Filtz L., Fu D.-O., Fuller W., Geringer C.,
 RA Gilmour R., Glaes J.S., Khoja H., Kraft A.R., Lagace R.E.,
 RA Leblanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
 RA Mcahren S.M., McHenry M., Mcleaster K., Mundy C.W., Niclas T.I.,
 RA Norris F.H., O'gara M., Peery R.B., Robertson G.T., Rocky P.,
 RA Sun P.M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
 RA Zook C.A., Baltz R.H., Jaskunas S.R., Kosteck P.R., Jr., Skatrud P.L.,
 RA Glass J.I.;
 RT "Genome of the bacterium Streptococcus pneumoniae strain R6."
 RL J. Bacteriol. 183:5709-5717(2001).
 DR EMBL; AE008479; AAK99864.1; -
 KM Complete proteome.
 SQ SEQUENCE 855 AA; 96177 MW; 4350E82A3F97089A CRC64;

Query Match 77.4%; Score 3224; DB 16; Length 855;
 Best Local Similarity 75.0%; Pred. No. 7.4e-176;
 Matches 617; Conservative 69; Mismatches 99; Indels 38; Gaps 4;

QY 1 SYELGLYQARTYKNNRVSITDGKATQKTEMLTPDEVSKREGINAEQIVIKITDQGYVT 60
 DB 32 AYEGLLQAOQTVKENNRVSITDGKATQKTEMLTPDEVSKREGINAEQIVIKITDQGYVT 91
 QY 61 SHGHHYHYNGKVPYDAIIEBELMKDPNYKLKDEDIVNEVGVYIVKDGKYVYLKDA 120
 DB 92 SHGHHYHYNGKVPYDAIIEBELMKDPNYKLKDEDIVNEVGVYIVKDGKYVYLKDA 151
 QY 121 AHAADVRTKEELNRKQEHSHQREGGTPRNDGAVALARASQGRYTTDDGYIFNADIIEDT 180
 DB 152 AHAADVRTKEELNRKQEHSHQREGGTPRNDGAVALARASQGRYTTDDGYIFNADIIEDT 211
 QY 181 GDAYIVPHGDHYHIFPKNELSASELAALAEFLSGGNLSNRTYTRQNSDNTSRNWPVS 240
 DB 212 GDAYIVPHGDHYHIFPKNELSASELAALAEFLSGGNLSNRTYTRQNSDNTSRNWPVS 271
 QY 241 VSNPCTNTNTSNNSNTNSQASQSDNDISLKLQYLKPLSQRHVSDDGIVPDAQITSRT 300
 DB 272 VSNPCTNTNTSNNSNTNSQASQSDNDISLKLQYLKPLSQRHVSDDGIVPDAQITSRT 331
 QY 301 ARGVAVPHGDHYHIFPKNELSASELAALAEFLSGGNLSNRTYTRQNSDNTSRNWPVS 360

Db 332 ANGVAAPHGDHFIPLYSQSLSPLEEKLAIIPLRYSNMWVPSRDEQSPSTPEPSPS 391
361 POPAPNLK-IDSN---SLVSQLVKVGEGYVEEKGISRYVFAKLPESTVANKLESKIS 416
Db 392 POPAPNPQAPSPNPIDEKLYEAVRKVGQYVEENGVPRIYIPAKLSMTAAGIDSKLA 451
Qy 417 KOESVSHLTJAKKENVAPRDOEFYDKAVNLTLEAHKALFXNKGKNSDFQALDKLERLND 476
Db 452 KOESLSHKLGAKTDLPSDSREFYNKAYDLARIHQDLDNKGRQYDFELDLNLERLND 511
Qy 477 ESTNKEKLVDDLLAFAPITTHPERLCKPNSQIEFYTEDVRIQALADKYTSDGYIFDEHD 536
Db 512 VSSDKKLVDLDLAFAPIRHPERLCKPNAQIYTTDEIOVAKLAKYTTEDGYIFDPD 571
Qy 537 IISDEGDVYTPHMGSHWIKGDSLSDKEVAAQAATKEKGLPSPADAVANPFGDGA 596
Db 572 IISDEGDVYTPHMTSHWIKKDSLEAEKAAQAAYAKEKGLTPSTHODSGNTEAKGA 631
Qy 597 AAIYNRKGEKRIPLVRLPYMVEHTVEVXNGNLIIPHKHYNHKEFAMFDDHTYKAPNGY 656
Db 632 EAIYNRVKAQKVPIDRMFYNLQYVEVNGSLIIPHYNHNIKFEWFDEGLYEAPEKGY 691
Qy 657 TLEDLFAITKYVEHDERPHSNDGKGNASEHVLGKXHSDEPNKFKADEE-----PV 710
Db 692 SLEDLFAITKYVEHDERPHSNDGKGNASDHVQRKNGQADTNQTEKENEKPEKDE 751
Qy 711 EET-----PAPEVPOVETKEVNAQLKEAVLLAKYT 742
Db 752 EETPEPEKQOSEKPSPKPTEPEPEESPESEPESEPEVETKEVKEKLEAEDLLGKIQ 811
Qy 743 DSSLKANATETLAGLRNNLTLOIMDNNSIAMEAEKLLALLKGS 785
Db 812 NIIKSNAKETLTGLKNLLFGTODNNTIAMEAEKLLALLKGS 854

RESULT 6
Q970M9 PRELIMINARY: PRT; 819 AA.

AC Q970M9: 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Conserved domain protein.
GN SP1174.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-TIGR4.
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., Deboy R.T., Haft D.H., Dodson R.D.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holzapfel E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
pneumoniae";
RT Science 293:498-506(2001).
RL EMBL: AB007418; AAK75283.1; -.
DR TIGR: SP1174; -.
DR InterPro: IPR006270; Strept_his_triad.
DR TIGRFAMs: TIGR01363; strep_his_triad; 2.
KW Complete proteome.
SQ SEQUENCE 819 AA; 92228 MW; 43852872E8163BDE CRC64;

Query Match 67.2%; Score 2797.5; DB 16; Length 819;
Best Local Similarity 67.0%; Pred. No. 1.6e-151;

Matches 546; Conservative 84; Mismatches 138; Indels 47; Gaps 7;
Qy 1 SYELGLQYA-RVKNKNNRVSYIDGQATQKNTLPDEVSKEEGINABQIVIKINDQGV 59
Db 21 SYELRQYQAGQKKSNNRVAYIDGQAGKABNLTPDEVSKKEGINABQIVIKINDQGV 80
Qy 60 TSHGDHYHYNGKVEYDAIISEELMKDPNFKLQKEDIYNEVKGQVIVKDGKYYVYLK 119
Db 81 TSHGDHYHYNGKVEYDAIISEELMKDPNFKLQKEDIYNEVKGQVIVKDGKYYVYLK 140
Qy 120 AAHADNVRTKEEINRQCKEHSQHREGGTPRNDGVALARSGQRYTTDGYIFNADIIED 179
Db 141 AAHADNVRTKEEIKQCKEHSQHREGGTPRNDGVALARSGQRYTTDGYIFNADIIED 197
Qy 180 TGDAYIVPHGDHYHIIPKNELSASGLAAEAALSGNLSNRTRRKQSDTSRTNWP 239
Db 198 TGDAYIVPHGDHYHIIPKNELSASGLAAEAALSGNLSNRTRRKQSDTSRTNWP 246
Qy 240 SVSNPGTTNTNTSNNSNQASQSDNDIDSLKQLYKLPLSORHYESDGLVPDPAQITSR 299
Db 247 NPAQRLSENHNLVTPTTHQ-NOGENTISLURELYAKPLSRHYESDGLIFDPAQITSR 305
Qy 300 TARGVAVPHGDHYHIIPIYSQSELEERARIIPLRYSNMWVPSRPEQSPQTPPEPSP 359
Db 306 TARGVAVPHGNHYHIIPIYQSELEERARIIPLRYSNMWVPSRPEPSPQTPPEPSP 365
Qy 360 GPQAPNLKIDSNSSLVQLVKVGEGYVEEKGISRYVFAKDLSEFTYKLNLESLSKQE 419
Db 366 SPQAPNSPFD-EKLVAEAVKVGQYVEENGVSRYIPADLSAETAGIDSKLAKQE 423
Qy 420 SVSHTLTJAKKENVAPRDOEFYDKAVNLTLEAHKALFXNKGKNSDFQALDKLERLNDST 479
Db 424 SLSHTLGTGRKTDLPSSDSREFYNKAYDLARIHQDLDNKGRQYDFELDLNLERLNDVS 483
Qy 480 NKEKLVDDLLAFAPITTHPERLCKPNSQIEFYTEDVRIQALADKYTSDGYIFDEHDIT 539
Db 484 DKYKLVEDLAFAPIRHPERLCKPNAQIYTTDEIOVAKLAKYTTEDGYIFDPRDITS 543
Qy 540 DEGDVYTPHMGSHWIKGDSLSDKEVAAQAATKEKGLPSPADAVANPFGDGA 599
Db 544 DEGDVYTPHMTSHWIKKDSLEAEKAAQAAYAKEKGLTPSTHODSGNTEAKGAEXI 603
Qy 600 YNRVAGEKRIPLVRLPYMVEHTVEVXNGNLIIPHKHYNHKEFAMFDDHTYKAPNGYTL 659
Db 604 YNRVGAQKVPIDRMFYNLQYVEVNGSLIIPHYNHNIKFEWFDEGLYEAPEKGYTL 663
Qy 660 DLFATIKYVEHDERPHSNDGKGNASEHVLGKXHSDEPNKFKADEE-----PVET 713
Db 664 DLFATIKYVEHDERPHSNDGKGNASDHVQRKNGQADTNQTEKENEKPEKDE 723
Qy 714 -----PAPEVPOVETKEVNAQLKEAVLLAKYTDSLSLKANA 750
Db 752 PREEKQSEKPSPKPTEPEPEESPESEPESEPEVETKEVKEKLEAEDLLGKIOPITKSA 783
Qy 751 TETLAGLRNNLTLOIMDNNSIAMEAEKLLALLKGS 785
Db 784 KETTLGLKNLLFGTODNNTIAMEAEKLLALLKGS 818

RESULT 7
Q9ANY3 PRELIMINARY: PRT; 819 AA.
AC Q9ANY3: 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Pneumococcal histidine triad protein B precursor (Fragment).
GN PHTB.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]

```

RP SEQUENCE FROM N.A.
RX MEDLINE=21101045; PubMed=11159990;
RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
RA Domitzier M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
RA Langerman S., Koenig S., Johnson S.;
RT "Identification and characterization of a novel family of pneumococcal
RT proteins (the phc family) that are protective against sepsis.";
RL Infect. Immun. 69:949-958(2001).
DR EMBL: AF318954; AAK06759.1; -
DR InterPro: IPR006270; strep_his_triad.
DR TIGRFAMs: TIGR01363; strep_his_triad. 2.
KW Signal.
FT SIGNAL
FT NON TER
SQ SEQUENCE 819 AA; 92108 MW; E602CFC16CC28A5F CRC64;

Query Match 66.9%; Score 2786.5; DB 2; Length 819;
Best Local Similarity 66.7%; Pred. No. 6.8e-151;
Matches 544; Conservative 84; Mismatches 140; Indels 47; Gaps 7;

OY 1 SYELGLVQA-RTVKENNRSYIDGKQATQKTENLTPPEVSKREGINAEQIVIKITDQGV 59
DB 21 SYELGRVAGQVKESNRVAYIDGQAGKAEMLTPPEVSKREGINAEQIVIKITDQGV 80
OY 60 TSHGDHYHYNGKVPYDAIISELLMKDPNPKLKDIEDIVNEVGGYVIKVDGKYVYLKD 119
DB 81 TSHGDHYHYNGKVPYDAIISELLMKDPNPKLKDIEDIVNEVGGYVIKVDGKYVYLKD 140
OY 120 AAHADNVRTKEINRQKQESHQREGGTPRNDGVALARSQGRYTTDGYIFNADITIED 179
DB 141 AAHADNVRTKEIKRQKQESHQREGGTPRNDGVALARSQGRYTTDGYIFNADITIED 197
OY 180 TGDVAVPHGHGHHYIIPKNEISASELAARAFSGRGLNSRTYRONSNDTSTWNPV 239
DB 198 TGDVAVPHGHGHHYIIPKNEISASELAARAFSGRGLNSRTYRONSNDTSTWNPV 246
OY 240 SVSNPGTNTNTSNNSTNSQASQSDNDISLLKQLYKLPLSRHVESDGLVFPDPAQITSR 299
DB 247 NPAQRLSENNLVTPTTHQ-NOGENISSLRRELYANPLSRHVESDGLVFPDPAQITSR 305
OY 300 TARGVAVPHGHGHHYIIPYSONSELEERTARIIPLYRSNHWVPSRPPSPQPTPEPSP 359
DB 306 TARGVAVPHGHGHHYIIPYSONSELEERTARIIPLYRSNHWVPSRPPSPQPTPEPSP 365
OY 360 GPOAPNLIKIDNSSLVQLVKGVEGFEKEGISRVPVAKDLSEYVKNLESLSKOE 419
DB 366 SPQAPSPNPID-GKLVEAKVAKGVDGVFEENGVSRYPYPAKDLSEYVAKGIDSKLAKOE 423
OY 420 SVSHTLTAKENAVAPRQDFYDKAVNLLTEAHKALFXKKGRNSDFQALDKLERLNDEST 479
DB 424 SLSHKLTGKTKTDLPESDREFFYKAYDLARIHQDLDDKKRGQVDFEALDNLERLKDVS 483
OY 480 NKEKLVDDLAFAPITPERIGKPNISOIETEDERVIAQLADKTTSDGTYFDEHDITIS 539
DB 484 DKVKLVEDILAFAPIRHPERIGKPNISOIETEDERVIAQLADKTTSDGTYFDEHDITIS 543
OY 540 DEGDVAVTPHGHGHHYIIPKNEISASELAARAFSGRGLNSRTYRONSNDTSTWNPV 599
DB 544 DEGDVAVTPHGHGHHYIIPKNEISASELAARAFSGRGLNSRTYRONSNDTSTWNPV 603
OY 600 YNRVAGEKRIPLVRLPYVWEHTVEVNGNLIIPHDKDHANIKFAFDDHTYKAPNGYLE 659
DB 604 YNRVAGEKRIPLVRLPYVWEHTVEVNGNLIIPHDKDHANIKFAFDDHTYKAPNGYLE 663
OY 660 DLPAITIKYVVEHDPHPSNDGNGVASEHVIGKQKSHSEDPNKNFPADEE-----PVEET 713
DB 664 DLPAITIKYVVEHDPHPSNDGNGVASEHVIGKQKSHSEDPNKNFPADEE-----PVEET 723
OY 714 -----PAPEVPOVETEKVLAQKAEVAVLAKVTDSSSLKANA 750
DB 724 PREKPOSEKPEPPTPEEPSEPSPESEPPVETEKVLEADLLGKIDPFIKANA 783
OY 751 TETLAGLNNLTLQIMDNNSIWAFAKLLALAKGS 785

```

```

DB 784 KETLTGLKNNLLFCTQNNNTIMAEAKLLALKES 818

RESULT 8
OYAN2 PRELIMINARY; PRT; 839 AA.
ID O9ANY2
AC O9ANY2;
DT 01-JUN-2001 (TREMUREL. 17, Created)
DT 01-JUN-2001 (TREMUREL. 17, Last sequence update)
DT 01-MAR-2003 (TREMUREL. 23, Last annotation update)
DE Pneumococcal histidine triad protein D precursor (Hypothetical protein
DE SP1003) (Fragment).
GN PHTD OR SP1003.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21101045; PubMed=11159990;
RA Adamou J.E., Heinrichs J.H., Erwin A.L., Walsh W., Gayle T.,
RA Domitzier M., Dagan R., Brewah Y.A., Barren P., Lathigra R.,
RA Langerman S., Koenig S., Johnson S.;
RT "Identification and characterization of a novel family of pneumococcal
RT proteins (the phc family) that are protective against sepsis.";
RL Infect. Immun. 69:949-958(2001).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eiseen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Duetkin A.S., Gwinn M., Kolony J.F., Nelson W.C., Peterson J.D.,
RA Umeyan L.A., White O., Salzberg S.L., Lewis M.R., Raskane D.,
RA Holtzapple E., Knouri H., Wolf A.M., Uetziack T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickenson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae.";
RL Science 293:498-506(2001).
DR EMBL: AF318955; AAK06760.1; -
DR EMBL: AE007403; AAK75120.1; -
DR TIGR: SP1003; -
DR InterPro: IPR006270; strep_his_triad.
DR TIGRFAMs: TIGR01363; strep_his_triad. 2.
KW Signal; Hypothetical protein; Complete proteome.
FT SIGNAL 1 29
FT NON TER 839 839
SQ SEQUENCE 839 AA; 93672 MW; 713B180D5E03BDCA CRC64;

Query Match 66.6%; Score 2772; DB 16; Length 839;
Best Local Similarity 65.0%; Pred. No. 4.7e-150;
Matches 541; Conservative 91; Mismatches 144; Indels 56; Gaps 8;

OY 1 SYELGLVQARV-KENNRSYIDGKQATQKTENLTPPEVSKREGINAEQIVIKITDQGV 59
DB 21 SYELGRVAGQVKESNRVAYIDGQAGKAEMLTPPEVSKREGINAEQIVIKITDQGV 80
OY 60 TSHGDHYHYNGKVPYDAIISELLMKDPNPKLKDIEDIVNEVGGYVIKVDGKYVYLKD 119
DB 81 TSHGDHYHYNGKVPYDAIISELLMKDPNPKLKDIEDIVNEVGGYVIKVDGKYVYLKD 140
OY 120 AAHADNVRTKEINRQKQESHQREGGTPRNDGVALARSQGRYTTDGYIFNADITIED 179
DB 141 AAHADNVRTKEIKRQKQESHQREGGTPRNDGVALARSQGRYTTDGYIFNADITIED 198
OY 180 TGDVAVPHGHGHHYIIPKNEISASELAARAFSGRGLNSRTYRONSNDTSTWNPV 239
DB 199 TGDVAVPHGHGHHYIIPKNEISASELAARAFSGRGLNSRTYRONSNDTSTWNPV 247
OY 240 SVSNPGTNTNTSNNSTNSQASQSDNDISLLKQLYKLPLSRHVESDGLVFPDPAQITSR 299

```

```

Db      248 NPAQRLSENHLVTPTTHQ--NÖGENISSLLRELYAKPLSERHVESDGLIFDPAQITSR 306
QY      300 TARGVAVPHGDHYHFIPIYSQMSLEERIAIIPLRYSNNHWVDSRPEQSPQTPPEPSP 359
Db      307 TARGVAVPHGDHYHFIPIYSQMSLEERIAIIPLRYSNNHWVDSRPEQSPQTPPEPSP 366
QY      360 GPQAPNLK-IDSN---SLVSQVLRKVGEGVFEFGKISRIVFAKDLPSFTVKNLSESL 415
Db      367 SPQAPNPQAPSNPIDEKLVKEAVRKVGVDGYFEENGVSRYIPADLSAETMAAGIDSKL 426
QY      416 SKQESVSHLTAKKENAVAPPDQEFYDKAYNLLTEAHKALFXNKGNSDFOALDKLLERLN 475
Db      427 AKQESVSHLTAKKKTPLPSDRREFYNNKAYDLARIHQDLDLNKRGQVDFEALDNLLERLK 486
QY      476 DESTNKEKLVDDLAFAPITHPERLGRKPSQIEYTEDEVRIAQLADKYTTSDGYIFDEH 535
Db      487 DVPSDKVKLVDDILAFAPIRPERLGRKPAQITTYDDEIÖVAKLAGKYTEDEGYIFDPR 546
QY      536 DIISDEGDAYVTPHMGSHWIGKDSLDEKVAQAQYTKKGLPPSPADYKAPNPTGDS 595
Db      547 DITSDEGDAYVTPHMGSHWIKDLSSEARAAQYAKKGLTPPSTDHQDSGNTGTEAG 606
QY      596 AAATYNRVKGKRIPLVLEPYWEHTVEYVNGNLIIPKHQYHNKIFAMFDHTTYAPNG 655
Db      607 AEAIVNRVAAKKVPLDRMPYNLQYTVENVNGSLIIPHYDHYNIKFEWFEDEGLYEAPKG 666
QY      656 YLLEDFATITKYVVEHPDERPHSNDGWNASEHVLGK-----KDHSE----- 697
Db      667 YLLEDFATITKYVVEHPDERPHSNDGWNASEHVLGK-----KDHSE----- 726
QY      698 -----DPNKNFKADEPVEETPAPEVPQVETEKYEAOLKEAEVLL 738
Db      727 ESDEKENHAGLNPSADNLKPSDTDETEBEAEDTTDEAEIIPQEVSVINAKADAEALL 786
QY      739 AAVTSSLKAAATETLAGLRNNLTLOIMDNNISMAEAKLLALLKGSNPSV 790
Db      787 EKVTDPISIRQNAEMETLTGLKSSLLTGKDNNTTISAEDSLLALLKESQAPPI 838

RESULT 9
QDQ008 PRELIMINARY; PRT; 853 AA.
AC QDQ008;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Pneumococcal histidine triad protein D.
GN PHTD OR SPR0907.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
CX NCBI_Taxid=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=2142945; PubMed=11544234;
RA Hoshino J., Alborn W.E. Jr., Arnold J., Blaszcak L.C., Burgett S.,
RA Dehoff B.S., Estrom S.T., Filtz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmore R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA LeBlanc D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA McAhren S.M., McHenry M., McMaster K., Mundy C.W., Niclas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rokey G.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Roestek P.R. Jr., Skatrud P.L.,
RA Glaes J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008464; AAK99711.1; -.
KW Complete proteome.
SQ SEQUENCE 853 AA; 95225 MM; 98F06A1ED390CF5 CRC64;

Query Match 65.6%; Score 2733; DB 16; Length 853;
Best Local Similarity 63.5%; Pred. No. 8.1e-148;

```

```

Matches 538; Conservative 92; Mismatches 145; Indels 72; Gaps 9;
QY      1 SYEIGLYOARTV-KENNRVSYIDGKOATOKETNLTPDEVSKREGINAQIVKITOGGV 59
Db      21 STELGRHQGVKKESNRVSYIDGQAKANLTPDEVSKREGINAQIVKITOGGV 80
QY      60 TSGDHYHYNNKVPYDAIISEELMKDPNYLTKEDIVNEVKGQVIVYDEKYYVYLYKD 119
Db      81 TSGDHYHYNNKVPYDAIISEELMKDPNYLTKEDIVNEVKGQVIVYDEKYYVYLYKD 140
QY      120 AAHADNVRTKEINRQKQESHQREGGTPRNDGVALARSQGRYTTDDGYIFNADIIED 179
Db      141 AAHADNVRTKEINRQKQESHQREGGTPRNDGVALARSQGRYTTDDGYIFNADIIED 197
QY      180 TGDAYIVPHGDHYHFIPIKNEISASLELAARLSGRGNLSNRTYRRQNSDNTSTNTVP 239
Db      198 TGDAYIVPHGDHYHFIPIKNEISASLELAARLSGRGNLSNRTYRRQNSDNTSTNTVP 246
QY      240 SVSNFGTTNTNTSNNSQASQGNIDISLLKQYKLPISQRYHVESDGLVFPDPAQITSR 299
Db      247 NPAQRLSENHLVTPTTHQ--NÖGENISSLLRELYAKPLSERHVESDGLIFDPAQITSR 305
QY      300 TARGVAVPHGDHYHFIPIYSQMSLEERIAIIPLRYSNNHWVDSRPEQSPQTPPEPSP 359
Db      306 TARGVAVPHGDHYHFIPIYSQMSLEERIAIIPLRYSNNHWVDSRPEQSPQTPPEPSP 365
QY      360 GPQAPNLK-IDSN---SLVSQVLRKVGEGVFEFGKISRIVFAKDLPSFTVKNLSESL 415
Db      366 SPQAPNPQAPSNPIDEKLVKEAVRKVGVDGYFEENGVSRYIPADLSAETMAAGIDSKL 425
QY      416 SKQESVSHLTAKKENAVAPPDQEFYDKAYNLLTEAHKALFXNKGNSDFOALDKLLERLN 475
Db      426 AKQESVSHLTAKKKTPLPSDRREFYNNKAYDLARIHQDLDLNKRGQVDFEALDNLLERLK 485
QY      476 DESTNKEKLVDDLAFAPITHPERLGRKPSQIEYTEDEVRIAQLADKYTTSDGYIFDEH 535
Db      486 DVSADKVKLVDDILAFAPIRPERLGRKPAQITTYDDEIÖVAKLAGKYTEDEGYIFDPR 545
QY      536 DIISDEGDAYVTPHMGSHWIGKDSLDEKVAQAQYTKKGLPPSPADYKAPNPTGDS 595
Db      546 DITSDEGDAYVTPHMGSHWIKDLSSEARAAQYAKKGLTPPSTDHQDSGNTGTEAG 605
QY      596 AAATYNRVKGKRIPLVLEPYWEHTVEYVNGNLIIPKHQYHNKIFAMFDHTTYAPNG 655
Db      606 AAATYNRVAAKKVPLDRMPYNLQYTVENVNGSLIIPHYDHYNIKFEWFEDEGLYEAPKG 665
QY      656 YLLEDFATITKYVVEHPDERPHSNDGWNASEHVLGK-----DHSDDPNK----- 701
Db      666 YLLEDFATITKYVVEHPDERPHSNDGWNASEHVLGK-----DHSDDPNK----- 725
QY      702 -----NFKAD-----EPPVETPAPEVPQVE 723
Db      726 EEDKEHDEVESEPTHEPSEDEKENHAGLNPSADNLKPSDTDETEBEAEDTTDEAEIIPQVE 785
QY      724 TEKVEAOKEAEVLLAKYTTDSGLKANATETLAGLRNNLTLOIMDNNISMAEAKLLALLK 783
Db      786 HSVINAKKAIEAPALAEKVTDSIRQNAVEITLTGLKSSLLTGKDNNTTISAEDSLLALLK 845
QY      784 GSNPSV 790
Db      846 ESQPPPI 852

RESULT 10
Q9ANY1 PRELIMINARY; PRT; 1039 AA.
AC Q9ANY1;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Pneumococcal histidine triad protein E precursor (Hypothetical protein SPI004).
GN PHTD OR SPI004.

```

[illegible]

QY	481	CEKVYDDLLAFLAPLPTHERERIGKPSQSEYETEDSVRIQIADKYTSSGYIFDEHDII	540
QY	411	QPTLPNNSLAPSP-SLPINPSTSEKHE-----EDYGFDAVRITAE	452
QY	541	EGDAVVTPEHMOGSHWIGKDSLSDKEKVAQAATYKEGILPPSPDAVDANPTGDSAAAIY	600
QY	453	DESGVMSHGHNHYHFFKKDLTEBQIKAAQKHLE-----VATSHNGDLSLSH	501
QY	601	NR-----VKGEKRIPLVRLPYMW--EHTVEKNGNLII-PKHQYNIKFAW	644
QY	502	EQDVPYNAKEMKMDLKKIEEKIAGIMQGYVAREEIVYNKEKMAIIYPHGHHHADP---	558
QY	645	FDDHTYKAPNGYLTLEDLFAITIKYYEHEDERPHSNDG-----WGNASEVILGCKHSDPN	700
QY	559	IDEH--KPVGIG---HSHSNYELFPEKEGVAKKEGKNKYTTGEBLTNVVNLKNSSTFNN	611
QY	701	KNF 703	
QY	612	QNF 614	
QY	241	VSNPCTTNTNTSNNSTNTNSQASQNSIDSLIKQLYKPLPSGRHVESDGLVDPDAQTSRT	300

```

Db      249 -----NFQVAKGSGTSPKANKSENLQSLKELYSPSAQRYSBDSGLVDFPAKIISTR 301
QY      301 ARGVAAPHGDHYHFIPIYSQWSELEERLARIIPLRYSNMHWVPDSRPEQSPQPTPEPSPG 360
Db      302 PNGVALPHGDHYFIPIYSKLSALEEKIARKVP----- 333
QY      361 POPAPNLKIDSNSLSVQLVRKVEGVPEEKGISRYVFAKDLPSFTVKLSLESKSKOES 420
Db      334 -----ISGTGSTVSTNAK-----PNEVSSLGSLSNPS 363
QY      421 VSHITLPAKKENVAAPROEFYDKAYNLITFAKALFPNKGKNSPQALDKLERLNDESTN 460
Db      364 ----LITSKELSSASDGYIFNPK-DIVEETATAYIYRHG--DHFIYIPK-----SNQIG 410
QY      481 KEKLVDLLFLAPLPIHPERLGRK-NSQIEYTEDEVRIQLADKYTTSDGYIFDEHDIIS 539
Db      411 QPTLPNNSLATPSP-----SLPIPDGISHEKHE-----EDGYGFDAIRIIL 451
QY      540 DEGDAYVTPHMGSHWIGKDSLSDKEKVAQAAYTKEGILPPSPDADVKANPTG-DSAAA 598
Db      452 EDSGFIIMSHGNHNFYFPPKDLTEBOIKAAQKHLE-----VKTSHGJLDSLS 500
QY      599 IYKRVKG-----EKRIPLVRLPYWV--EHTVEYKGNLII-PKHDIYHNIKFA 643
Db      501 HEODYPGNAKEMKDLPKLIEEKIAGIMKQYGVKRESIVVAKENKALIIYPHGDHHAHP-- 558
QY      644 WPDHHTYKAPNGVLTLEDLFAITIKYYVHPDERPHSNDG-----WGMSSEHTLGGKDSSEDP 699
Db      559 -IDH--KPVGIG--HSHSNVELFKPEEGVAKKEGKRVYTGEBELTVNVLKNTSTW 610
QY      700 NKNF 703
Db      611 NQNF 614

RESULT 12
ID 099XV4 PRELIMINARY; PRT; 825 AA.
AC 099XV4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein spv2006.
GN SPY2006.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
RA MEDLINE=21192684; PubMed=1126296;
RA Ferretti J.J., McGhan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Szecze S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R., Song L., White J.,
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663 (2001).
DR EMBL; AB006623; AAK34688.1;
DR InterPro; IPR006270; Strep_his_triad.
DR TIGRFAMs; TIGR01363; strep_his_triad; 4.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 825 AA; 92649 MW; 57DF50969BD50F4 CRC64;

Query Match 22.8%; Score 949; DB 16; Length 825;
Best local Similarity 29.5%; Pred. No. 5.9e-46;
Matches 277; Conservative 100; Mismatches 214; Indels 348; Gaps 34;
QY 1 SYELGILYQARTVENNRVSTID--GKQATQKTEMLTPDEVSKREGIMAEQIYIKITDQG 57
Db 22 SYOLGHHMGSAATKQIAYIDSKGAKAPKT-NKTMQDISAEIGSISAEQIVVKITDQG 80

```

```

QY 58 YVTHGDHYHYNGKVPYDAIISEELMKDPNYLKDEDIIVNEKGVYIKVDGKYVYL 117
Db 81 YVTHGDHYHYNGKVPYDAIISEELMTDPYRFQSDVINEILDGVIKNGVYVYL 140
QY 118 KQAARADVNRTEELINRQKQHSQH-REGG-----TPRNDGAYALANSQGRYTTDDG 169
Db 141 KQSRKKNIRITQOIAEOVAKGTKEAKKEGLAQVAHLSEEVAAVNEAKRQGRYTTDDG 200
QY 170 IFNASDIEDTGDAYIVPHGDHYHYIPKNELSASEVLAABEVL--GRGNLSNSTRYR 226
Db 201 IFSPTDIDDLGDAYIVPHGNHYHYIPKDLSPSELAAQAAVWSQKQNG--ARSDYRP 258
QY 227 QNSDNTSRTNW-VPSVS-NPGTNTNTSN-----NSNTNSQASQSD-----IDSLL 271
Db 259 TPAPAPGRKAPIPDVTWPGQGHQDNGCYHAPRPRDAQNKHQPREFGKPTKELL 318
QY 272 KQYKPLPSQRYVESDGLVFPDQITSRARAVAPHGCHYFIPIYSQWSELEERLARI 331
Db 319 DQLRDLKYRHYVEEGILFEBPTQVAKSNAPGVVPHGDHYHYIPRQSLPEMELAD-- 376
QY 332 PLRYSNMHWVPDSRPEQSPQPTPEPSPGPQAPNLKIDSNSLSVQLVRKVEGVPE 391
Db 377 --RYLAGQTEDD-----DSGSD----- 391
QY 392 KGISRYVFAKDLPSFTVKNLBESKLSKQESVSHTLPAKKENVAAPROEFYDKAYNLITFAH 451
Db 392 -----HSPSDKE-VHTIFLGR-----IKAY----- 412
QY 452 KALFNKGNNSPQALDKLERLNDESTNKEKLVDDLALFLAPITHPERLPGNSQIEYT 511
Db 413 -----GKGLD-----GKP----- 420
QY 512 EDEVRIQLADKYTTSDGYIFDEHDIISDEGDAYVTPHMGSHWIGKDSLSDKEKVAQA 571
Db 421 -----YDSDAYVFESKISHSVKSGVTAKRGDHPHYIGFELBOYEYDEVAN 468
QY 572 YTKEGILPPSPDADVKANPTGDSAAAIYRNVKGEKRIPLV-----RLPYNV 618
Db 469 WKAKG-----QADELAALDQGGKEK-PLPPTKYSRKVTYKDGKGYVM 513
QY 619 -----EH-----TVEYKN-----GN 628
Db 514 PKDGKDYFARQDLDTQIAFAEQEIMLKDKKHRYDVTDTGIEPRLADVSSLPMHAGN 573
QY 629 -----LIIPKHDIYHNIKFAWPDHHTYKAPNGVLTLEDLFAITIKYYVHPDERPHSND 680
Db 574 ATYDGGSFVPIPHIDIHVVPYSWL-----TRDQIATIKYVWQHEVVR--D 618
QY 681 GW-----GNASEVLGKDHSEDPNKNFKADEPVEETPAE-----PE----- 718
Db 619 IYKSGHESSGVIRVVTPLDGRAGMFPWQIHSABEYQKALABGRFATFPDGYITPDPDV 678
QY 719 -----VPOVE-----TEKYAOLKEAEVLLAKVTSLSLANATETLA 755
Db 679 LAKEFPVWKDGSFSPRADGSLRLTKNSDLSQAEOQAOELLAKGN-----AGDAYDT-- 732
QY 756 GLRNLTLQIMNNSIMAEKLLALKSNSSYSKEX 794
Db 733 -----DKPREKQQADK--SNENQPSSEASKEX 757

RESULT 13
ID 08NZ82 PRELIMINARY; PRT; 823 AA.
AC 08NZ82;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hypothetical protein spvM18_2072.
GN spvM18_2072.
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.

```



```

Db      419 YDTSADVRSK-----ESTHSVD-----KGVAKHGHDHNYIGFGELEQY 459
QY      451 HKALPYN-----KGRNSDFQALDKLRLNDESTNKEKLVDDLLAFLAPITHEPRLG----- 502
Db      460 ELDEVANWVKAKGQ-----ADELAAL--DQEGCKRPLEDTTKVSRKTKKCKGVYMP 512
QY      503 KNSQIETYTEDEVRIAQLADKYYTSDGYIFDEHDIISDEGDVYTPHMGSHWIGKDSLS 562
Db      513 KKGKDYFYARDQDLDTLQIA-----FAEQELM-----LK 540
QY      563 DEXKAAQAQYTKKGLPPSPDADYKANTGDSAAIYRVKCKRILPLVRLPYMEHYV 622
Db      541 DKHRYRDIY--DTGEP-----RLAVDSSSLPMAGNAT 573
QY      623 EVKNGNLIIPKDHYNIKFAPWDHTYKAPNGYTLEDFATIKYVEHPDERPHSDGM 682
Db      574 YVTGSSFYVPHIDHIVPYSWL-----TRDQATYKVMQIEVPR--DW 618
QY      683 -----GNASEHVLGKKDHSEDPKNKFADEPEVEETPAE-----PE----- 718
Db      619 SKPGHESSGVIPNVTPLDKRAGMPWQIHSABEVQKALAEGRFATPDGYIFDPEDVLA 678
QY      719 -----VPOVE-----TEKVEAQLKEAEVLLAKVTDSSLKANATETLAEL 757
Db      679 KETFWKDGSSFSIPRADGSSLRTINKSDLSQAEMWQAQBELAKN-----TGADTDT-- 730
QY      758 RNNLTQLQIMDNNSIMAEKLLALLKGSNPSSVSXK 794
Db      731 -----DKPEKQQAOK--SNEHQPSSEASKEE 755

```

RESULT 15

```

093GT5      PRELIMINARY;      PRT;      825 AA.
ID 093GT5;
AC 093GT5;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Histidine triad protein of group A streptococci.
GN HTPA.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RP SEQUENCE FROM N.A.
RC STRAIN=SSI-9;
RA Teroo Y., Kawabata S., Hamada S.;
RT "Characterization of a novel histidine triad protein of group A
RT streptococci."
RT Submitted (NOV-2001) to the EMBL/Genbank/DBJ databases.
RL EMBL; AB073859; BAB71774.1;
DR InterPro; IPR006270; Strept_his_triad.
DR TrifPfam; TIGR01363; strep_his_triad.4.
SQ SEQUENCE 825 AA; 92623 MW; DEAECC199181DFB CRC64;

```

Query Match 22.7%; Score 944; DB 2; Length 825;

Best Local Similarity 29.4%; Pred. No. 1,1e-45; Mismatches 214; Indels 348; Gaps 34;

```

QY      1 SYELGLYQARTVKNRNVSYID--GKQATOKTENTLPPEVSKREGINAEOIVIKITPDG 57
Db      22 SYQLGKHMGSAATKONQIAYIDSKKAKAPKT--NKTMOQIASBESISAEQIVAKITPDG 80
QY      58 YVTSGDHHYVNGKVPYDAIISSEILMKDPYKLDKEDI VNEVGKGVYIKVQKYYVL 117
Db      81 YVTSGDHHYVNGKVPYDAIISSEILMTDPMYRFKQSDVINIILDGVIYIKVNGVYYVL 140
QY      118 KLAADNADNRTEKEINRQKQHSQH--REG-----TRNCGAVVLAASQRYTTDDXY 169
Db      141 KGSGRKNRRTKQQLAEQVAKGKKEKGLAQVAHLKSEEVAAVNAEAKRQGRYTTDDY 200
QY      170 IFNASDIIDTDGDAYIVPHGDHYVLPKNELSASEILAEEAFIS--GRNLSNSRTYR 226

```

```

Db      201 IFSPDIIIDLDGADALVPHGNHYHYIPKDKLSPESSLAQAQVWSQKQGRG--ARPSDYR 258
QY      227 QNSDNTSRTNM--VPSVS--NPGTTNNTNS-----NSGTNSQAQSDN-----IDSL 271
Db      259 TPAPAGRRKADIPVTPNPGQGHQPDNGGYHAPAPRPDPAQONKHQRDEFKGTFFKEL 318
QY      272 KQYKLPISQRHVESDGLVFPDPAQITSRFARGVAPHGHHYFIPYSQSELEERLARI 331
Db      319 DQHLRLDKRYRVEEDGLIFETQYIKSNAFGVYPHGDHIIITRSQSLPEMELAD-- 376
QY      332 PLVRSNMHWPDSPREQSPQPTPEPSPQOPAPNLKIDSNSLVSQLVKRVGEGYFEE 391
Db      377 --RYLAGQTEED-----DSGSD----- 391
QY      392 KGISRVPAKDLPSFTVKNLESKLSKQESVHTLTAKENAVAPROEFDKAYNLITEAH 451
Db      392 -----HSKPSDKK--VYTHPLGHR-----IKAY----- 412
QY      452 KALFYKNGRNSDFQALDKLRLNDESTNKEKLVDDLLAFLAPITHEPRLGKPNQIEXT 511
Db      413 -----GKGD-----KRP----- 420
QY      512 EDEVRIAQLADKYYTSDGYIFDEHDIISDEGDVYTPHMGSHWIGKDSLDPKEVAAQA 571
Db      421 -----YDTSADVRSKESIHVSVDKSGVAKGHDHNYIGFGELEQYELDEVAN 468
QY      572 YTKEGILPPSPDADYKANTGDSAAIYRVKCKRILPLV-----RLPYV 618
Db      469 WYKAKG-----QADELAALDQEGKER--PLPDTKVSARKVTGDKGVYWM 513
QY      619 -----EH-----TVEYKN-----GN 628
Db      514 PKDGQDYIARDQDLDTQIAPEQELMKDKKHRYVDVDTGIEPRLAVDSSSLPMAGN 573
QY      629 -----LIIPKDHYNIKFAPWDHTYKAPNGYTLEDFATIKYVEHPDERPHSD 680
Db      574 ATYDGSFVPIPHIDHIVPYSWL-----TRDQATIKYVMQIEVPR--D 618
QY      681 GW-----GNASEHVLGKKDHSEDPKNKFADEPEVEETPAE-----PE----- 718
Db      619 IWSKPGHESSGVIPNVTPLDKRAGMPWQIHSABEVQKALAEGRFATPDGYIFDPEDV 678
QY      719 -----VPOVE-----TEKVEAQLKEAEVLLAKVTDSSLKANATETLA 755
Db      679 LAKETFWKDGSSFSIPRADGSSLRTINKSDLSQAEMWQAQBELAKN-----AGADTDT-- 732
QY      756 GLRNNLTQLQIMDNNSIMAEKLLALLKGSNPSSVSXK 794
Db      733 -----DKPEKQQAOK--SNEHQPSSEASKEE 757

```

RESULT 16

```

08K501      PRELIMINARY;      PRT;      823 AA.
ID 08K501;
AC 08K501;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Histidine triad protein.
GN SPW3 1724.
OS Streptococcus pyogenes (serotype M3).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=198466;
RP SEQUENCE FROM N.A.
RC STRAIN=MGAS315 / Serotype M3;
RA MEDLINE=22133808; PubMed=12122206;
RA Bares S.B., Sylva G.V., Barbican K.D., Lei B., Hoff J.S.,
RA Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
RA Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
RA Schlievert P.M., Musser J.M.;

```

"Genome sequence of a serotype M3 strain of group A Streptococcus:

RT phage-encoded toxins, the high-virulence phenotype, and clone

RT emergence."

Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083 (2002).

DR EMBL; AE014169; AAM8031.1; -

DR InterPro; IPR006270; Strep_his_criad.

DR TrEMBL; TIGR01363; strep_his_criad; 1.

KM Complete proteome.

SO SEQUENCE 823 AA; 92431 MW; 8CC6BD517A2DD616 CRC64;

Query Match 22.6%; Score 940; DB 16; Length 823;

Best Local Similarity 30.0%; Pred. No. 1.9e-45;

Matches 262; Conservative 122; Mismatches 272; Indels 216; Gaps 30;

```
1 SYELGLYQARTVKNRNVSYID--GKQATOKTENLTPDEVSKREGINAEQIVIKITDOG 57
22 SYQLGKHHMGLATKQNOIAYIDSKGKAKAPKT-NKTMDOISAEBSISAEQIVVKITDOG 80
58 VYTSHGDIHYHYNGKVPYDAIISBELLMKDPNYKLDEDIVNEVKGVIKVDGKYVYL 117
81 VYTSHGDIHYHYNGKVPYDAIISBELLMTPDNYHKQSDVINELDGVIVKNGYVYI 140
118 KDAHAADVRTKEINRQKQHSQH-REGG-----TPRNDGAVALARSGRYTTDDGY 169
141 KPGSKRKNIRTKQOIAEQVAKGTKEAKGGLAQVAHLKEEVAANAEAKRQGRYTTDDGY 200
170 IFNSADIIEDTGDATVPHGDHYHYIPKNEISASFLAAAEAFSL--GRGNLSNRTYR 226
201 IFSPTDIIDLDGDAVLPHGNNHYHYIPKDDLSPELAAQAWYSQKQGRG--ARPSDYP 258
227 QNSDNTSRITMWPVS-VS-NPGTTNTNTSN-----NSNTSQASQSDN-----IDSLLK 272
259 TPAQRARRAP-IPVDTNPGQGHQPDNGNGYHAPRRPNDASQNKQRQREPFKTKFELLD 317
273 QLYKLPISQHVESDGLVFPDAQITSRTARGVAVPHGDHYHYIPYSQMSLEERITARIIP 332
318 QLMRLDLKYRHVEEDGLIFEPYQVIXSNAFGYVAVPHGDHYHYIIPRSQSLPEMELAD-- 374
333 LRYRNMHVWPDSPRQPSQPPTPEPSPQAPAPNLKIDSNSLSVQLVRKVGEGYVEEK 392
375 -RYLAGQ-----TEDNSGSDHSSKSDKEVTHTFLHRKAKGKGLDGPRY 419
393 GIS-RYVFAKDLPSSTVKNLSKLSKQESVSHTLTAKKENVAPRQDEFYDKAYNLTEAH 451
420 DTSDAVYFSK-----ESHSVD-----KSGVYAKHDHNYIGFGLBQYE 460
452 KALFKN---KGRNSDFQALDKLRLNDESTNKEKLVDDLLAFLAPITPBERLGPNSQ 507
461 LDEVANWVYAKQO-----ADELAALDQEQGKEKPLFD----- 493
508 IEYFDEVRIAQOLAKYTTSDGYITDEHDIISDEBDAYVTPHMGSHMIGKDSLDEKV 567
494 -----TKVSRKVTQDGK-----GIMPRQDGYVYARQDL--L 527
568 AQAQYTKKEGILPSPADAVKANPTGDSAAAIYNKVEKRIPLVRLRYMEHTYEVKNG 627
528 TQIAAEDELMKDKNHYRYDIVDQ-----IEPRLADVSSLPMHAGNATYYDGS 578
628 NLIIPEKHQYHNKIFAWFDHTYKAPNGVYLEDLFATIKYUYEHPDERPHSNDG----- 682
579 SFVIRPHIHIVVYPSWL-----TRDQATIKYVWQHDEVP--DVMSKPGH 623
663 ---GNASHEVILKQHSDEDPNKPFAKDEPVEETAFA-----PE----- 718
624 EESGSVINVPRLDRAGMFMWQIISAEVQKALAEGRFATPDGYIFDPRDVLAKETV 683
719 -----VPOVE-----TEKVEAOLKEAEVLAKVTDSSLKANATETLAGRNLT 762
664 WKDGFSFIPRADGSLRTINKSDLSQAEVQAOELLAKN-----AGDATDT----- 730
763 LQIMDNNSIMAEKLLALLKGSNPSVSKEK 794
731 ----DKPRKQOQADK--SNEHQOSEASKEE 755
```

RESULT 17

Q8DZ81

ID Q8DZ81

AC Q8DZ81

DT 01-MAR-2003 (TREMBlrel. 23, Created)

DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)

DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)

DE Streptococcal histidine triad family protein.

GN SAG1233.

OS Streptococcus agalactiae (serotype V).

OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;

OC Streptococcus.

OX NCBI_TaxID=216466;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=2603 V/R / Serotype V;

RX MEDLINE=2222988; PubMed=12200547;

RA Tetteijn H., Masignani V., Cieslewicz M.J., Eisen J.A., Peterson S.,

RA Wessels M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,

RA Masoff L.C., Wolf A.M., Beaman M.J., Brinkac L.M., Daugherty S.C.,

RA DeBoy R.T., Durkin A.S., Kolonay J.F., Madupu R., Lewis M.R.,

RA Radune D., Fedorova N.B., Scanlan D., Khouli H., Mulligan S.,

RA Carcy H.A., Cline R.T., Van Aken S.E., Gill J., Scarselli M., Mora M.,

RA Jacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Matone D.,

RA Rinaldo D., Rappuoli R., Telford J.L., Kasper D.L., Grandi G.,

RA Fraser C.M.;

RT "Complete genome sequence and comparative genomic analysis of an

RT emerging human pathogen, serotype V Streptococcus agalactiae."

RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396 (2002).

DR EMBL; AE014248; AAN0011.1; -

DR TIGR; SAG1233; -

KM Complete proteome.

SQ SEQUENCE 822 AA; 92400 MW; 8CC6BD516727F98 CRC64;

Query Match 22.5%; Score 937; DB 16; Length 822;

Best Local Similarity 29.5%; Pred. No. 2.9e-45;

Matches 272; Conservative 98; Mismatches 225; Indels 318; Gaps 32;

```
1 SYELGLYQARTVKNRNVSYID--GKQATOKTENLTPDEVSKREGINAEQIVIKITDOG 57
22 SYQLGKHHMGLATKQNOIAYIDSKGKAKAPKT-NKTMDOISAEBSISAEQIVVKITDOG 80
58 VYTSHGDIHYHYNGKVPYDAIISBELLMKDPNYKLDEDIVNEVKGVIKVDGKYVYL 117
81 VYTSHGDIHYHYNGKVPYDAIISBELLMTPDNYHKQSDVINELDGVIVKNGYVYI 140
118 KDAHAADVRTKEINRQKQHSQH-REGG-----TPRNDGAVALARSGRYTTDDGY 169
141 KPGSKRKNIRTKQOIAEQVAKGTKEAKGGLAQVAHLKEEVAANAEAKRQGRYTTDDGY 200
170 IFNSADIIEDTGDATVPHGDHYHYIPKNEISASFLAAAEAFSL--GRGNLSNRTYR 226
201 IFSPTDIIDLDGDAVLPHGNNHYHYIPKDDLSPELAAQAWYSQKQGRG--ARPSDYP 258
227 QNSDNTSRITMWPVS-VS-NPGTTNTNTSN-----NSNTSQASQSDN-----IDSLLK 272
259 TPAQRARRAP-IPVDTNPGQGHQPDNGGYHAPRRPNDASQNKQRQREPFKTKFELLD 317
273 QLYKLPISQHVESDGLVFPDAQITSRTARGVAVPHGDHYHYIPYSQMSLEERITARIIP 332
318 QLMRLDLKYRHVEEDGLIFEPYQVIXSNAFGYVAVPHGDHYHYIIPRSQSLPEMELAD-- 374
333 LRYRNMHVWPDSPRQPSQPPTPEPSPQAPAPNLKIDSNSLSVQLVRKVGEGYVEEK 392
375 -RYLAG-----QTDNDNS-----GSDH----- 390
393 GISRYVFAKDLPSSTVKNLSKLSKQESVSHTLTAKKENVAPRQDEFYDKAYNLTEAH 452
391 -----SKPSDK-VHTTFLGHR-----IKAY----- 410
453 ALFYKNGNSDFQALDKLRLNDESTNKEKLVDDLLAFLAPITPBERLGPNSQIEYTE 512
```

```

Db      411  -----GKGLD-----GKP-----418
QY      513  DEVRIAQLADKYTTSNGYIFDEHDIISDEGDAYVTPHMGSHWIGKDSLDEKKAQAAY 572
Db      419  -----YDTSAYVFSKESHSVDKSGVTAKGHDHFHYIGFGELOYELDEVANW 467
QY      573  TREKG---ILPPSPDADVKANPTGDSAAA-----LYNR-----602
Db      468  VKAKGQADELVAAALDOEGKEKPLPDTKVKSRKVTGDKGVYIMPKGKDYFYARQOLD 527
QY      603  -----VKGEKR-----IPVLRPYMVEHTVEVKNGLIIPKH 636
Db      528  TQIAFAEOELMKDKGHYRDIYDTGIEPRLAVDSLPMHAGNATYDGSFVIPHIDH 587
QY      637  YHNIRKAFDDHHTYKAPNGTILEDLFATIKYVVEHDERPHSNDG-----GNASEH 688
Db      588  IHVPRYSWL-----TRNQIATIKYVQHPEVRP---DVMSKPGHESSGSVITP 632
QY      689  VLGKDHSEDPNKNFKADEEPEETPAE-----PE-----VP 720
Db      633  VTPLDKRAQMPNQIHSABEVOKALABGRFAAPDGYIPDRVYLAKETFWKDGFSFIP 692
QY      721  QVE-----TEKVEAQLKEAEVLLAKVTDSSLKANATETLAGLRNNLTLOIMDNNSI 771
Db      693  RADGSSLRTINKSDLSQAEWQQAQELLAKKN---AGDATDT-----DKPEE 735
QY      772  MAEAEKLALLKGSNPPSSVSKER 794
Db      736  KOQADK---SNENQOPSEASKER 755

```

RESULT 18

```

Q9ZHG7  PRELIMINARY; PRT; 822 AA.
AC      Q9ZHG7;
DT      01-MAY-1999 (TrEMBLrel. 10, Created)
DT      01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT      01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE      Hypothetical 92.4 kDa protein.
OS      Streptococcus agalactiae.
OC      Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC      Streptococcus.
NCBI_Taxid=1311;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=R268;
RA      MEDLINE=99115568; PubMed=9916102;
RA      Spellerberg B., Rozdzinski E., Martin S., Weber-Heymann J.,
RA      Schultze N., Lueticken R., Podbielski A.;
RT      "Lmb, a protein with similarities to the Lrai adhesin family, mediates
RT      attachment of Streptococcus agalactiae to human laminin."
RL      Infect. Immun. 67:871-878(1999).
DR      EMBL; AF062533; AAD13797.1;
DR      InterPro; IPR006270; Strept_his_triad.
DR      TIGRFBAS; TIGR01363; strep_his_triad; 4.
KW      Hypothetical protein
SQ      SEQUENCE 822 AA; 92386 MW; 80E4BD913481F98 CRC64;

```

Query Match 22.4%; Score 934; DB 2; Length 822;
 Best Local Similarity 29.4%; Pred. No. 4.2e-45;
 Matches 271; Conservative 99; Mismatches 235; Indels 318; Gaps 32;

```

QY      1  STELGIYQARTVKENNRVSYID--GKQATQKTENTLPDESVKREGINAEQIVIKTTDQ 57
Db      22  SYTLGHHGMLATKDNQIAVIDSKGVKAPKT-NKTMQDISAEEGISAQIVVKTLDQ 80
QY      58  YVTSHGDIHYVYVYKPYDAIIESELLMKDQPNKLTDEDIYVNVKGVYIKDQGVYVY 117
Db      81  YVTSHGDIHYVYVYKPYDAIIESELLMTDQPNYHFKQSDVITNITLQGVYIKVGNYYVY 140
QY      118  KQAAHADVNTKEINROKQESHQ--REGG-----TPRNDGAVVALARSQGRVYTTDDGY 169

```

```

Db      141  KPSRRKNIPTKQOIAEQAQKGTKEKGLAQVAHLSKEEVAAVNEAKRQRYTTDDGY 200
QY      170  IFNASDIEDTGDAYIVPHGDHYHYI PKNELSASELAABEAFLS---GRGLNSNRTYR 226
Db      201  IFSPTDIIDLDGAYIVPHGNHYHYI PKKDLSPSELAAQAQWQKQKRG--ARSDYRP 258
QY      227  QNSDNTSRITNWPSVS-NPGTNTNTSN-----NSNTNSQAQSQND-----IDSLIK 272
Db      259  TPAPGRKAP-IPDVTNPGQGHQDPDNGGVYHAPRPNDASQNKQRDFFKKTFFELLD 317
QY      273  QLYKPLSORHYVESDGLVDVPAQITSRTARGAVPHGDHYHYI PYSQMSLEBRARIIP 332
Db      318  QLRDLKRYHVEEGLIFEPQVIXSNAFGVVPHGDHYHYI PRSQSLPLEMLAD--- 374
QY      333  LRRSNHWVPDRPQRPQPTPEPSPGQAPAPNLKIDNSSLVQLVARKNVEGYFEER 392
Db      375  RYLLAG-----QTDNDS-----QSDH-----390
QY      393  GISRYVFAKDLPEETVKNLESKLSKQESVSHLTAKENVAAPRDOEFDYKAYNLLTEAKH 452
Db      391  -----SKPSDKE-VTHFFLGHR-----IKAY-----410
QY      453  ALFXNKGNSDPQALDKLIERLNDSTNKEKLVLDLAFATITPERLQKNSQIETYE 512
Db      411  -----GKGLD-----GKP-----418
QY      513  DEVRIAQLADKYTTSNGYIFDEHDIISDEGDAYVTPHMGSHWIGKDSLDEKKAQAAY 572
Db      419  -----YDTSDAYVFSKESHSVDKSGVTAKGHDHFHYIGFGELOYELDEVANW 467
QY      573  TREKG---ILPPSPDADVKANPTGDSAAA-----LYNR-----602
Db      468  VKAKGQADELVAAALDOEGKEKPLPDTKVKSRKVTGDKGVYIMPKGKDYFYARQOLD 527
QY      603  -----VKGEKR-----IPVLRPYMVEHTVEVKNGLIIPKH 636
Db      528  TQIAFAEOELMKDKGHYRDIYDTGIEPRLAVDSLPMHAGNATYDGSFVIPHIDH 587
QY      637  YHNIRKAFDDHHTYKAPNGTILEDLFATIKYVVEHDERPHSNDG-----GNASEH 688
Db      588  IHVPRYSWL-----TRNQIATIKYVQHPEVRP---DVMSKPGHESSGSVITP 632
QY      689  VLGKDHSEDPNKNFKADEEPEETPAE-----PE-----VP 720
Db      633  VTPLDKRAQMPNQIHSABEVOKALABGRFAAPDGYIPDRVYLAKETFWKDGFSFIP 692
QY      721  QVE-----TEKVEAQLKEAEVLLAKVTDSSLKANATETLAGLRNNLTLOIMDNNSI 771
Db      693  RADGSSLRTINKSDLSQAEWQQAQELLAKKN---AGDATDT-----DKPEE 735
QY      772  MAEAEKLALLKGSNPPSSVSKER 794
Db      736  KOQADK---SNENQOPSEASKER 755

```

RESULT 19

```

Q9E338  PRELIMINARY; PRT; 481 AA.
AC      Q9E338;
DT      01-MAR-2003 (TrEMBLrel. 23, Created)
DT      01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT      01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE      Hypothetical protein.
OS      Streptococcus agalactiae (serotype III).
OC      Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC      Streptococcus.
NCBI_Taxid=216495;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=NEM316 / Serotype III;
RA      MEDLINE=22342508; PubMed=12354221;
RA      Glaser P., Rusniok C., Buchrieser C., Chevalier F., Frangoul L.,

```



```

OY 236 MWVPSVNPQTNTNTSNNSTNSQASQSDNDISLLKQYLPLSQRHVESDGLVEPDAQ 295
DB 206 -----GIMHTSDGFLFNGG 221
OY 236 IITSRTARGVAVPHGDHYHFIPIYSQSELEERLARIIPLYRSNHWVPSRPPQSPQPP 355
DB 222 IKGTPPTGLVVEHHNHLHFISF--ADLRKGGWGIADRYQPOKKV-DSKQSPS----- 272
OY 356 EPSPGQAPAPNLKIDNSSSLVSQLVRYK-----GE-GYVFEKGISRYVFA 400
DB 273 --SKPRPTENTLPKDKL-KYLARELHLDISRIKVLKTLNGEIGFEYPHDHTHVIA 329
OY 401 KDLPSRTVKNLESKLSKQSVSHITLAKKENVAPRQEFYDYKAYNLTFAHKLFXNKR 460
DB 330 KDI-----DLSKPIPNH-----HDEDHKHHHDESDHK----- 360
OY 461 NSDFQALDKLERLNDSTNKEKLVDD--LAFAPITHPERLQKPSQIEYTEDVRI 517
DB 361 -----HEEHETKSNKLSDEQKLIYLA-----EKLGLNPNQIEV----- 396
OY 518 AGLADKYTSDGYIPDEHDIISDEGAYVTPHMGSHW-----IGK-----DLSLD 563
DB 397 -----LTSEDGSIIFKY-----PHDHSHTIASKDIEIGKIPDGHHDHSHA 438
OY 564 KEKVAQAAYTKKKG-----IL--PSPDADVKNP--TGDSAAALYNRVKGKRIIP 610
DB 439 KKKV-GMATLKQIGFDEIIODLIHADAPTPPSNETNEPKRQMLATYTKINIGRTNP 497
OY 611 LVRLPYVMEHTEVKNKGNLIIPHKD-----HYNHIKFAW-----FDDHTY--KAP----- 653
DB 498 FQRFGLSLMPNIEVL-GIGFTPIINDMTPLYQFKKQLKQMLMTNTGITDYSFLDKMPLLEG 556
OY 654 -----NGYTLDEDFAITKY-----YVENP-----DERPFS 678
DB 557 DISQNG--IKDLSFLTKYKQLSLIAAANGITSLKPLAELPNLOFLVLSHNNISDLTPLS 614
OY 679 N-----DGMGNASEHVLGKKDSE--DPNKNFKADEEPVEETPAE----- 716
DB 615 NITKQELHLDHNNVKNLSA-LSGKKDLKVLDSNNKKSADLSTLKTSTLETLLETNTS 673
OY 717 -----PEVPOV-----ETEKVEAQLKEAEVLLAKVTDSLK----- 747
DB 674 NLSFLKQNPVKNLITINNAKLASLDGIESEDEIVKVEAGNQIKSLVLKNGKSLKFLNV 733
OY 748 -ANAETIAGLRNNLTLOIMDNNSIMAEKALILKGSNPSSVSEKI 795
DB 734 TNNQLTSLLEGVNNYTSLETL-----SVSKNKL 760

RESULT 22
OQE029 PRELIMINARY; PRT; 877 AA.
AC Q8E029;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Protein of unknown function/lipoprotein, putative.
GN SAG0907.
OC Streptococcus agalactiae (serotype V).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_Taxid=216466;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=2603 V/R / Serotype V;
RX MEDLINE=22222988; PubMed=12200547;
RA Terelein H., Masiagnati V., Cieslewicz M.J., Eisen J.A., Peterson S.,
RA Wesele M.R., Paulsen I.T., Nelson K.E., Margarit I., Read T.D.,
RA Madoff L.T., Wolf A.M., Beaman M.U., Brinkac L.M., Daugherty S.C.,
RA DeBoy R.T., Doult A.S., Kolonay J.F., Madupu R., Lewis M.R.,
RA Radune D., Fedorova N.B., Scanlan D., Khouiri H., Mulligan S.,
RA Carly H.A., Cline R.T., Van Aken S.E., Gail J., Scarselli M., Mora M.,

```

```

RA Iacobini E.T., Brettoni C., Galli G., Mariani M., Vegni F., Malone D.,
RA Rinaldo D., Rappapoli R., Telford J.L., Kasper D.L., Grandi G.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative genomic analysis of an
RT emerging human pathogen, serotype V Streptococcus agalactiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:12391-12396 (2002).
DR EMBL; AB014233; AM99793.1; -.
DR TIGR; SAG0907; -.
KW Lipoprotein; Complete proteome.
SQ SEQUENCE 877 AA, 97563 MW, 4829E9E53A4F27 CRC64;

Query Match 7.3%; Score 302; DB 16; Length 877;
Best Local Similarity 20.8%; Pred. No. 5.8e-09;
Matches 200; Conservative 95; Mismatches 263; Indels 402; Gaps 43;

OY 13 KENNNAVSTIDGKOATOKTENLTPEV---SKREI-----MAEQIYIKTD 55
DB 26 QEHKNSHHIKTKQVAKKANKKVKSVKSHKKRKAGVDFPTDDGFLITKDSKILSHTD 85
OY 56 OGVTSHGDHYHYVNGKVPYDAIISELLMKDPNKLDEDIYNKGVYIKVDGKYV 115
DB 86 SCIVVAHGNSHFL-----FYSDLGSKFS 110
OY 116 YLKDAAHADNVATKEINROKQESHQREGTTPRNDGAVALARSGRYTTDDGYIFNASD 175
DB 111 YL-----IPNANTKTNKQAVRNFKA-----GAVAV-----NTLNDGYFNPAD 149
OY 176 IIEDTGDAIYVPHGDHYHFIPIKNELISASELAAAEFLSGRGLNSRTRYRONSNTSRT 235
DB 150 IVEEDANGYVVRHGHGFYIPKASLSQOKQVQASRAVSRIGQSNSS--HYRNNSSKIAGL 207
OY 236 MWVPSVNPQTNTNTSNNSTNSQASQSDNDISLLKQYLPLSQRHVESDGLVEPDAQ 295
DB 208 HY-PT-----SNGFLFNGG 221
OY 236 IITSRTARGVAVPHGDHYHFIPIYSQSELEERLARIIPLYRSNHW-----VPS 344
DB 222 IKGTPPTGLVVEHHNHLHFISPADL-----RKGGSIADRYQPOKKADS 266
OY 345 RPEQSPQTPPEPSGQAPAPNLKIDNSSSLVSQLVRYK-----GE-GYVF 389
DB 267 KKQSPS-----SKPRPTENTLPKDIKDKL-KYLARELHLDISRIKVLKTLNGEIGFEY 318
OY 390 BEKGSRYVPAKDLSEVTVKNLESLSKQSVSHITLTKKENVAPRQEFYKAVNLTLE 449
DB 319 PHDHTHVIMAKDI-----DLSKPIPNH-----HDEDHKHHHDES 357
OY 450 AAKALFXNKNRNSDFQALDKLERLNDSTNKEKLVDD--LAFAPITHPERLQKPS 506
DB 358 DHK-----HEEHETKSNKLSDEQKLIYLA-----EKLGLNPN 392
OY 507 QIEYTEDEVRJAQLADKYTSDGYIFDEHDIISDEGAYVTPHMGSHW-----IGK- 558
DB 393 QIEV-----LTSEDGSIIFKY-----PHDHSHTIASKDIEIGKIP 427
OY 559 -----DLSDEKVAQAAYTKKKG-----IL--PSPDADVKNP--TGDSAAAI 599
DB 428 IPDGHHDHSHADKV-GMATLKQIGFDEIIODLIHADAPTPPSNETNEPKRQMLAV 486
OY 600 YNRVKGKRIPLVRLPYVMEHTEVKNKGNLIIPHKD-----HYNHIKFAW-----FDDHT 649
DB 487 TKINIGQNTNPFQRLGLSLMPNIEVL-GIGFTPIINDMTPLYQFKLQKMLMTNTGITDYS 545
OY 650 Y--KAP-----NGYTLDEDFAITKY-----YVENP----- 672
DB 546 FLDKMPLLEGLDISQNG--IKDLSFLTKYKQLSLIAAANGITSLKPLAELPNLOFLVLS 603
OY 673 -----DEBPHN-----DGMGNASEHVLGKKDSE--DPNKNFKADEEPVEETPA 715
DB 604 HNNISDLTPLSLTKQELVLDHNNVKNLSA-LSGKKDLKVLDSNNKKSADLSTLKTSTSL 662
OY 716 E-----PEVPOV-----ETEKVEAQLKEAEVLLAK 740

```

```

Db      663 ETLNNTNTSFLKONPKVSNLTINNAKLASLDGIEESDEIVKEAGNQIKSLVLK 722
Oy      741 VTDSSLK-----ANATETLAGLRNNULTQIMDNNSIMAEKLLALKSGNSPSVSEKI 795
Db      723 NKQSLKFLVNTNNQLTSLGVNNTSLETL-----SVSKNKL 760

RESULT 23
O8D006 PRELIMINARY: PRT: 182 AA.
ID O8D006
AC O8D006:
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Pneumococcal histidine triad protein E, truncation.
GN PHNE-TUNCACTION OR SPR0910.
OS Streptococcus pneumoniae (strain ATCC BAA-255 / R6).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=171101;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21429245; PubMed=11544234;
RA Hoskins J., Alborn W.E. Jr., Arnold J., Blaszczak L.C., Burgett S.,
RA Dehoff B.S., Estrem S.T., Filtz L., Fu D.-J., Fuller W., Geringer C.,
RA Gilmour R., Glass J.S., Khoja H., Kraft A.R., Lagace R.E.,
RA Leahen D.J., Lee L.N., Lefkowitz E.J., Lu J., Matsushima P.,
RA Mcahren S.M., McInerney M., Mcleaster K., Mundy C.W., Nicas T.I.,
RA Norris F.H., O'Gara M., Peery R.B., Robertson G.T., Rockey P.,
RA Sun P.-M., Winkler M.E., Yang Y., Young-Bellido M., Zhao G.,
RA Zook C.A., Baltz R.H., Jaskunas S.R., Rostock P.R. Jr., Skatrud P.L.,
RA Glass J.I.;
RT "Genome of the bacterium Streptococcus pneumoniae strain R6.";
RL J. Bacteriol. 183:5709-5717(2001).
DR EMBL; AE008464; AAK9714.1; -.
KW Complete proteome.
SQ SEQUENCE 182 AA; 20145 MW; 29E12D653113A43 CRC64;

Query Match 5.8%; Score 243; DB 16; Length 182;
Best Local Similarity 31.6%; Pred. No. 1.5e-06;
Matches 61; Conservative 16; Mismatches 50; Indels 66; Gaps 5;

Oy      147 TPRNDGAVALARSGRYTTDDGYIFNADIEDTGDAIVPHGDHYHYIPKNELSASELA 206
Db      4 TPNN-----GVSAVDGCVENPNDIVRDGTDAIVRHGDHYHYIPKSLN----- 48
Oy      207 AAEAFLSGRGULSNRTRYRQNSDNTSRTNVVPSVSNFGTTTNTSNNSTNSQASQSDND 266
Db      49 -----NPSHSNTEEEVSSSSS 65
Oy      267 IDSLKQLYKLPLSQRHVESDGLVFPDPAQITSRARGVAVPHGDHYHYIPYSQMSLEER 326
Db      66 V-----LSNPSLHVHHEEDSDGHFDANKRIISEDEGFIIPHGDHNYHYIKV-QTGYEAA 118
Oy      327 IARIIPLRYSNH 339
Db      119 LKNKIP-SLQSNV 130

RESULT 24
O99276 PRELIMINARY: PRT: 792 AA.
ID O99276
AC O99276:
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Putative internalin A precursor.
GN INLA OR SPI1361.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / Serotype M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Perrett J.J., Moshan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Stvorov A.N., Kenton S., Lai H.S., Lin J.P.,
RA Qian Y., Jia H.G., Najaf F.Z., Ren Q., Zhu H., Song J., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AE006574; AAK34188.1; -.
DR InterPro; IPR001611; LRR.
DR InterPro; IPR007092; LRR_SDS22.
DR Pfam; PF00560; LRR; 4.
DR PROSITE; PS50504; LRR_SDS22; 1.
KW Complete proteome.
SQ SEQUENCE 792 AA; 87458 MW; 9D5E32288485ACE0 CRC64;

Query Match 5.4%; Score 223; DB 16; Length 792;
Best Local Similarity 19.5%; Pred. No. 0.00016;
Matches 168; Conservative 116; Mismatches 279; Indels 300; Gaps 42;

Oy      1 SYELGVQARTVKENNRYSYIDGKQATOKT-ENLTPEVSKREG--INAEQIVKITPOG 57
Db      30 TYPKTKQSRKGMTSNKIKPKKSKTKTKHKGAVGVDFPTDDGFLTKDSKILSKTDQG 89
Oy      58 VYTSGHGDHYH-----YNGKVPYDAIISSELT-----MKDP--NYKDKD 94
Db      90 IVVDHGDHSRIFPAADLKGFPEYLIRKGLASLAPAVAGRAASGTSKVADPHHHYENP 149
Oy      95 EDIVNEVKGVIKVDGKYVYLKDAHADNVRTKEINRQKQSHQRECGTPRNDGAV 154
Db      150 ADIVAEADALGYTVARHDFHYILKSS-----LSGQTQAAQVATRLPOTSILV 198
Oy      155 ALARSQG---RYTDDGYTFNADIEDTGDAIVPHGDHYHYIPKNELSASELAAEA 210
Db      199 STATANGIPGLHFPISDGFQFGQIGVYKDSILVVDGHLHPSFADL----- 248
Oy      211 FLSGRGLNSRTRYRQNSDNTSRTNW--VPSVSNFGTTTNTSNNSTNSQASQSDND 268
Db      249 -----RQC-----GMAHVADQYDPAKKAAPRTHOTPLPSSEKRYQ 286
Oy      269 SLKQLYKLPLSQRHVESDGLVFPDPAQI---TSTARGAVAPHGHDHYHYIPYSQMSLE 324
Db      287 EKLAYL-----AEKLGIDPSTIKRVETDQGLGLEYRPHDHAHV---MLSDIE 332
Oy      325 ERIARIIPLRYSNHVWDSRP-----EOPSPQPTPE 356
Db      333 -IGKDIP---DPALEHARELEKHVKVMDTLRALGDFDEVILDIYRTHDAPTPPSNE 386
Oy      357 PSPGPOP---APNLKIDSNS-----SLVSQLVRKVGEGY----- 387
Db      387 KDPMMKEMKVLATVTKLDGSRKDPLOKRGSLNLNL-ETLIGTGPIMDISPLVQFKLK 445
Oy      388 -VFEKGISRYVPAKDLP-----SEYKNELESKISQESVSHYTAKEN---VA 433
Db      446 QLMTKGTGVTDYRFLDNNPQLEGIDISQNNKDI-SFLSKYKNL--TLVAADNGIEDIR 502
Oy      434 PRDGFYDKANLILTEAHKALFXNKGKNSDQALDKL--LERLNDSE-----TKKE 482
Db      503 PLGQ-----LPNL-----KFLVLSNNKISDLSPLASLHQLOELHIDNNQITDLSPVSHKE 552
Oy      483 KL--VD-----DLAFLAP-----ITHPELT-GKPN----- 505
Db      553 SLTVVDSLRNADVLAITQAPKLETLAMVNDTKVSHDLFLKNPNPLSSISINRAQLQSLEG 612
Oy      506 -----SQLEYTEDEVARIQADK-----YTTSDGY-IPDEHDIISDEGDA 544
Db      613 IEASSVIRVVEAEGNQKSLVLKDKQGLFLFDVTGNQLTSLGVNNTFALDILSVSKNQ 672
Oy      545 VYTHMGH-----SHWIGKDSLSD---KKVAAQATYRKGLIPSPDADVAKNP 591
Db      673 LTNVNLSPKNTVTNIDISH--NNISLADKLKLNQHOIPEALAKN--FPAYVEGSMVNGG 727

```

QY 592 TGDSSAAIYNRKGEKRIPLVRLPYVEHTEVKNGLIIPKDHYNIKFAWFDHTYK 651
 DB 728 TAEKKAAMATKAKESAOEASESHDYNHNTYEDDEGHA-----HEHROK-----DDHDH- 776
 QY 652 APNGYTLLEDFAIKYKVEHPDE 674
 DB 777 -----EHEDE 781

RESULT 25

Q8K714 PRELIMINARY; PRT; 792 AA.
 AC Q8K714;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Putative internalin A.
 GN INLA OR SPYM3_1035.
 OS Streptococcus pyogenes (serotype M3).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=198466;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MGAS315 / serotype M3;
 RA MEDLINE=22133808; PubMed=12122206;
 RA Beres S.B., Sylva G.L., Barbican K.D., lei B., Hoff J.S.,
 Mammarella N.D., Liu M.-Y., Smoot J.C., Porcella S.F., Parkins L.D.,
 Campbell D.S., Smith T.M., McCormick J.K., Leung D.Y.M.,
 Schlievert P.M., Musser J.M.;
 RA "Genome sequence of a serotype M3 strain of group A Streptococcus:
 RT phage-encoded toxins, the high-virulence phenotype, and clone
 RT emergence".
 RL Proc. Natl. Acad. Sci. U.S.A. 99:10078-10083(2002).
 DR EMBL: AE014157; AAM79642.1; --
 DR InterPro: IPR007092; LRR_SDS22; 1.
 DR PROSITE: PS50504; LRR_SDS22; 1.
 KW Complete proteome; Hypothetical protein.
 SQ SEQUENCE 792 AA; 87476 MW; 1D501C439CAE224B CRC64;

Query Match 5.1%; Score 213; DB 16; Length 792;
 Best Local Similarity 19.9%; Pred. No. 0.006;
 Matches 120; Conservative 91; Mismatches 211; Indels 182; Gaps 27;

QY 1 SYELGIYQARTVENNRVSYIDGKATQKT-ENLTDEYSKREG--INAEQIYIKITDQG 57
 DB 30 TYPITKQSRKGMTNKNIKPIKSKTKNTKNGAVDEPTDGFILTKDSKILSKTDG 89
 QY 58 YVTSQDHYHY-----YNGKVPYDAIISEELL-----MKDP--NYTKD 94
 DB 90 IYVDHGHGHFIIFYADLKSPFEYILPKGASLAKPAVAQOAAOQSKVAADPHNHYEFPN 149
 QY 95 EDIVNEVGKGYVIKVDGKYVYLKDAHADNVTKKEINROKQESHQREGGTPRNDGAV 154
 DB 150 ADIVAEDALGYTRHDHDFHYILKSS-----LSGOTQAKQOVATRLPQSSPV 198
 QY 155 ALARSG-----RTTDDGYFNASDIIDTGDAYIYPHGDNHYIIRKNEISABELLAER 210
 DB 199 STRTAGITGLHFPYTDGFOFNGOGIVGVKDNILVDHGHLPISPADL----- 248
 QY 211 PLSGRNLNSRTYRQNSDNTSRTNW--VPSVSNCGTTNTNTSNNSNTNSQASQNDID 268
 DB 249 -----RQG-----GMAYVADQYDPAKKAKEKAPENHHTPELSEREKEVQ 286
 QY 269 SLIKQILYKPLSQRHVESDGLVFPDPAQI--TSRTARGVAVPHGDHYHFIYQSOMSELE 324
 DB 287 EKLAIV-----AEKLGIDPSTIKKVEFQDGLGLEVPHDHAHL--MLSDIE 332
 QY 325 ERIARIILPKRNSHWVPSRPF-----EQSSPQTEB 356
 DB 333 --IGKIP--DPHAIIEHARELEKHKFGMDTLRALGFDEEVLIDIVRTHDAETPPSPNE 386
 QY 357 PSEGPDP--APMLKIDSNS-----SLVSGLVKRGEGY----- 387

DB 387 KDPNNMKEWLAVIKLIDLSRKPLOKGLSILPNL-ETLIGGFTPIKDISPVLQFKUX 445
 QY 388 --VFEKGISRYFADLP-----SETVKNLEKSKQESVSTTLPAKKENVAPRQ 437
 DB 446 QLLMTKTGVTDYRFIDNMPQLGIDISQNNLKDI-SFLSKYKYL--TLVAADN-GIEPI 501
 QY 438 EFYDKAVNLLTEAHKALPYNKGKNSDPQALDKLERINDESTNKEKLVLDLAFLPIFH 497
 DB 502 RSLGQLPNL-----KFLVLSNNKIDSLPLAS-LHQLGLHIDNNQITD-----LSPVSH 550
 QY 498 PERL 501
 DB 551 KESL 554

RESULT 26

Q963T1 PRELIMINARY; PRT; 1078 AA.
 AC Q963T1;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Glutamate-rich protein (Fragment).
 GN GLURP.
 OS Plasmodium reichenowi.
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=5854;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=2133688; PubMed=11420113;
 RA Theisen M., Thomas A.W., Jensen S.;
 RT "Cloning, nucleotide sequencing and analysis of the gene encoding the
 RT glutamate-rich protein (GLURP) from Plasmodium reichenowi".
 RL Mol. Biochem. Parasitol. 115:269-273(2001).
 DR EMBL: AF356828; AAK40236.1; --
 FT NON TER 1
 FT NON TER 1
 FT NON TER 1078
 SQ SEQUENCE 1078 AA; 123581 MW; 8D388D88B23913C CRC64;

Query Match 4.7%; Score 197.5; DB 5; Length 1078;
 Best Local Similarity 19.6%; Pred. No. 0.0072;
 Matches 164; Conservative 136; Mismatches 347; Indels 191; Gaps 37;

QY 7 YQARTYKENNRVSYIDGKATQKTENLTP--DEVSKREGINAEQIVIKITDQGV--- 59
 DB 372 HNVNVLQENNNINNHOLEPQEKANIESFEKNIIDSEIILPENVEKEIIVDVSPKHFHNE 431
 QY 60 -----TSHGDHYHYNGKVPYDAIISEELMKDPNPKLDEDIVNEVGKGYVIKVDGKY 114
 DB 432 TLEQETSESEHEAVSEKNAHETVHEHEAVSOGESNPEKADND--GNVSQNSNNELNENF 489
 QY 115 VYLKDAH--ADNVRT-----KEEI--NRQKQESHQ--HREGTPRNDGAVALAR 159
 DB 490 VESKSEHEPAPNEBESLEAHQELVPEQNQOSESGLVNDGEGFEAHQBELVPEQ 549
 QY 160 QGRYTTDDGYINNASDIIDTGDAYIVPHGDHYH---YIPKELASLAAAEALASR 215
 DB 550 NNOESGESKLVNDGEGFEAHQELVPEQNQOSESGLVNDGEGFEAHHEHNSSEE 609
 QY 216 GNLSNRTYRQNSDNT-----SRTWVPSVSNPGTTNTNTSNNSNTNSQASQNDID 270
 DB 610 SN-SELHEHEVESDSDPEPPEHEHVEVSEENPEASENEBESIEAHQELIESQNDSE 668
 QY 271 LKQLYKPLSQRHVESDGLVFPDPAQI--TSRTARGVAVPHGDHYHFIYQSOMSELEBRIAR 329
 DB 669 LNE-----NELVSEKSVSEPAHEVILVSEKSVSE--AEHVEIYSEKSVSEPAHVES 720
 QY 330 IIPKRYRSHWVPSRPFQSPQTPPEPS--PGQAPANL-----KIDNSSLVS-QLYR 381
 DB 721 VSE-RGASE--PSEHVESVSEQSNNEPEKKDGVPSPFEEIEKVDVQPKIVDQIIE 776
 QY 382 KYGEGYVFEKGISRYVFAKDLPSFTVKLJESKLQESVSHTLTKAKKENVAPRQDEPYD 441

```

Db 777 -----PWFVLDQPNPQEPVPSFVKIEKVPSEEN-----KKGSDVPEVEE--- 816
Oy 442 KAYNLTTEAHKALFKXKGRNSDFQALDKLERLNDESTNKEVLVDLLAFLAPITHPERL 501
Db 817 -----KENVSEVVEKEKNPOELFBEI-----PLKDE-I 844
Oy 502 GKPNQIETDEDEVRIAQLADKYTSDGYIFDEHDIIISDEGDAYVTPHMGSHWIGKDSL 561
Db 845 EKIEBELF--EDVTEQLDLEHKTVDPEIIEVEI-----PSLHENEVAHEI 892
Oy 562 SDEKVAQAAYTKEGILPPSPDAD--VKANPTGDSAAAIYNRKGEKRIPLVRLPYW 618
Db 893 VEIEEVF-----PEPNQNEPEPEINEDDKSAHIQHEIYVEVEIILPEEDKNEKV 940
Oy 619 EH-TVEVKNGLIIPKDHYNHFKAFWPDHDTYKAPNGTYLEDLFAITIKYYEHDEDEPH 677
Db 941 EHEIYVEE--ILPEENKNEKV-----EHE-----IYVEEILPE 973
Oy 678 SNDGNGNASEH-VLGGKDHSEDPNKNPKFADEEPEVETPAEPPOVETKEVQAOLKEAV 736
Db 974 ENE--NEKVEHEIYVEEILPEEDKNEKVEHEIYVEEILPE--ENKKEKHEIYVEE 1029
Oy 737 LIAKVTDSLKANATETTLAENLTLQIMDNNSIMAEAKLLALKGSNPSVSEK 794
Db 1030 ILPEIIE-----IEVPSQTNNEENIE-----TIKPEKK-----NEFSVVEEK 1068

```

RESULT 27

```

O77033 PRELIMINARY; PRT; 1390 AA.
AC 077033;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE TREA.
GN TREA.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
CX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=AX2;
RA Saito J., Adachi H., Sutoh K.;
RT "Dictyostelium TREA homologous to yeast Ssn6 is required for normal
RL growth and early development."
DR EMBL; AB009080; BAA3143.1; -
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 9.
DR SMART; SMO0028; TPR; 9.
SQ SEQUENCE 1390 AA; 160421 MW; BAAB992656002DE CRC64;

```

Query Match 4.6%; Score 192.5; DB 5; Length 1390;
 Best Local Similarity 18.1%; Pred. No. 0.02;
 Matches 147; Conservative 131; Mismatches 291; Indels 245; Gaps 35;

```

Oy 15 NNRYSYIDGKATQCTEULTPEVSKREGSINAQIVITIDOGYVTSGDHYHYNGKVP 74
Db 690 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 736
Oy 75 YDAIISSELLMKDPYKDKEDIVNEVKGVIYK-----DGKYVYLKDAADHNVRT 128
Db 737 -----DNVNSKNDVLDRLRYKGLIEREKTPNGDGR-----DNRDNIKD 775
Oy 129 KEINRQKQSHQREGGTPR-----NDGAVALA-----RSQGRYTTDDG 168
Db 776 NRD-NRDRGRDRNDRGDSRDRIOEYTRVNNNNNNNNNNSSINNMMNNNNNNNNNN 834
Oy 169 YIFNASDIETDGAIVP---HGDHYIYIPKNEISASELAAEAFLSGRGLNSRTYR 225
Db 835 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 885

```

```

Oy 226 RQNSDNTSRTNWPVSU-----SNPGTTNTNTNSNTNSQASQSDID-SLLKQL 274
Db 886 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 941
Oy 275 YKLPLSQRHVESDGLVPD-PAQITSRT-ARGVAVPHGDHYHFIFYSONSELEERAIIP 332
Db 942 ALSFOSSQHKDRRIIIDEESDINERSKTRPSI-----VKAEKEKREYIV 988
Oy 333 LRVSNHWPVPSRREQSPQPTPEPSPGQAPAPLKI-DSNSLSVOLVRVGVGYVEE 391
Db 989 DKERS-----PTPIITTEKPEKQVETKIDKSSLVKEVDKE-----NEK 1027
Oy 392 KGISRYVPAKDLPEETVGNLESKLSKQSVSHTLTAKKENVAPRDOEFYDKANLITTEAH 451
Db 1028 ESPSSSSSSKIEKEETEKEKEKEKEKEVEKEIENDKEKE-----KEVE 1078
Oy 452 KALFXNKGNSDFQALDKLER-----LNDESTNKEVLVDLLAFLAPITHPERLGP 505
Db 1079 KDVEENK--SVEKSEKPEVEKESTTTTNDDEGE-----LSEPTTTTKD 1123
Oy 506 SQIETDEEVRIAQLADKYTSDGYIFDEHDIIISDEGDAYVTPHMGSHWIGKDSLDPKE 565
Db 1124 DSKLPTDEKLLSSVSPPTTAVEO-----SRDETELE 1156
Oy 566 KVAQAAYTKEGILPPSPDADVKANPTGDSAAA-IYNRKGEKRIPLVRLPYWHEITVEV 624
Db 1157 MDTKEDSEK-----KSTTTTAAASSVAPRIDEKKSPTT-----TTT 1198
Oy 625 KGNLTIIP-HKDHYNHFKAFWPDHT-----YKAPNGTYLEDLFAITIKYYVEHPD 673
Db 1199 TMTTVEPTTHDKSSSKN---DTTTTTTTTTTTAKSPVSSPTRS-----D 1242
Oy 674 E--RPHSNDGNGNASEHVLGGKDHSEDPNKNPKFADEEPEVETPA-----EPVPOV 722
Db 1243 EVEPHQ---DASQEBINRKLEDDITSTPSRLPDPSTSSATTAITSBPQESP-L 1296
Oy 723 ETEKVEAQLKEAVLAKVTDSSILKANATETTLAAG 756
Db 1297 KKENPVGETLSPEIKDKSSSSSSSSSSSSSSSINTG 1330

```

RESULT 28

```

O8P0G5 PRELIMINARY; PRT; 792 AA.
ID O8P0G5;
AC O8P0G5;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Putative intercalin A.
GN SPYMI8.1373.
OS Streptococcus pyogenes (serotype M18).
OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
OC Streptococcus.
CX NCBI_TaxID=186103;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=MGS8232 / Serotype M18;
RA MEDLINE=21927593; Pubmed=11917108;
RA Smoot J.C., Bardhan K.D., Van Gompel J.J., Smoot L.M., Chaussee M.S.,
RA Syuva G.L., Sturdevant D.E., Ricklets S.M., Porcella S.F.,
RA Parkins L.D., Beres S.B., Campbell D.S., Smith T.M., Zhang Q.,
RA Kapur V., Daly J.A., Vasey L.G., Musser J.M.;
RT "Genome sequence and comparative microarray analysis of serotype M18
RT group A Streptococcus strains associated with acute rheumatic fever
RT outbreaks."
RL Proc. Natl. Acad. Sci. U.S.A. 99:4668-4673(2002).
DR EMBL; AE010057; AAL97968.1; -
DR InterPro; IPR001611; LRR.
DR InterPro; IPR007092; LRR_SDS22.
DR Pfam; PF00560; LRR; 4.
DR PROSITE; PS50504; LRR_SDS22; 1.
DR Complete proteome; Hypothetical protein.
SQ SEQUENCE 792 AA; 87440 MW; SC4DDEBEI8A91A7 CRC64;

```


Db	4907	-YSDSDSI	SYEHF	FDNGSESD	DDIR	RYEEEM	DDDYME	EEEBI	VNO	DGE	EEDEE	FA	4966							
Qy	408	VKNLSKLSK	QESVSH	TLTAKEN	VAPR	DOEY	-----	-----	-----	-----	-----	DKAYNL	448							
Db	4966	IEG	-EDDEA	QDEEN	DEEGN	ENETL	FNDEY	YNTD	NNND	NNND	NNND	NNNN	502							
Qy	449	EAHKA	LYFNK	RNSD	FPAL	DKLER	LAND	EST	YNKE	KLVD	DLA	FLAP	508							
Db	5025	DEENN	MSNN	DEEN	DERDE	DDVE	DDDE	DVE	-----	VDD	-----	DEVD	507							
Qy	509	EYTED	EVN	IA	LAKY	TTSD	QI	PE	PHI	-----	-----	ISDE	551							
Db	5073	DVDEE	EEVEE	EMD	MENED	-	DEDD	I	DDDD	DEDE	DEDE	DEDE	512							
Qy	552	HSW	IGK	SL	SDKE	VAA	QAAT	KE	GL	IPSP	AD	VKAN	611							
Db	5125	-----	DEVD	DE	EV	LD	DE	-----	DE	DI	DE	-----	516							
Qy	612	VR	LP	WE	HT	VE	-----	KN	GL	II	PK	-----	648							
Db	5165	VH	-AY	V	QDR	NEI	DE	QI	YNS	N	DL	LPHE	522							
Qy	649	TYK	APNG	TT	LE	D	PAT	I	KY	YEH	DER	HS	708							
Db	5224	DM	EQ	N	V	R	IE	IE	ET	-----	EE	DE	527							
Qy	709	PVE	ST	PA	PE	VQ	VE	T	KE	V	EA	Q	764							
Db	5277	DE	DE	DE	DE	DE	GE	EE	Y	DE	DE	DE	533							
Qy	765	IM	DNS	IM	772	-----	-----	-----	-----	-----	-----	-----	765							
Db	5335	INV	NN	SL	5342	-----	-----	-----	-----	-----	-----	-----	5335							
RESULT 33																				
ID	Q25860	PRELIMINARY;	PRT, 1271 AA.																	
AC	Q25860																			
DT	01-NOV-1996	(TREM	BL	REL	01	Created)														
DT	01-NOV-1996	(TREM	BL	REL	01	Last sequence update)														
DT	01-OCT-2002	(TREM	BL	REL	22	Last annotation update)														
DE	Glutamate rich protein.																			
GN	GLUP.																			
OS	Plasmodium falciparum.																			
CC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.																			
OX	NCBI_TaxID=56833;																			
RN	[1]																			
RP	SEQUENCE FROM N.A.																			
RX	MEDLINE=92131041;	PubMed=1775153;																		
RA	Borre M.B., Dziegiel M., Hogh B., Petersen E., Riiseck K., Riley E.,																			
RA	Meis J.F., Aikawa M., Nakamura K., Harada M., Wind A., Jakobsen P.H.,																			
RA	Cowland J., Jepson S., Axelisen N.H., Vuust J.;																			
RT	"Primary structure and localization of a conserved immunogenic																			
RT	Plasmodium falciparum glutamate rich protein (GLUP) expressed																			

```

OY 115 VYLKDAAHADN-VRTKEEINROKQHSQREGGTPRNDGAVALLARSQGRYTTDDGYTINA 173
D 478 -----SQSNNEELNENEFVESSEKSEH-----AAENESSLEECH---H 513
OY 174 SDII-----EDTGDAYIVPHGDHYHYPKNEIASELAAEAFLSGRGLNSRTYRQ- 227
D 514 EIIVEEQNNEESGSEKLV-----DNDGCFEEAHHEHFS---EVSSEINENEF 560
OY 228 -NSDSTSTNNVPSVSNPCTTNTNTSNNSTNSQASQSDNDISLLKOLYKPLSGRHVES 286
D 561 VESDK-----SVTEPAHEEVSESNPEPAENESSSIEEAOE--EIVPEQNDSE 610
OY 287 --DGLV-----FDDPAQITS-----RTARGVAVPHGHHYHPIYSQMS 322
D 611 GESGVDNEBEGDFEERPHNEEPDPNDSELSNELVESSEKSYSEF-AEHVEIVSEKSYSE 669
OY 323 LEEBRIARIIPLRYSNNHWVPSRPEQSPQPTPEPS---PGQAPANL---KIDSNSL 375
D 670 PAEHV-EIVSEKSTSE---PAEHVESVSESNNEPESEKKGDPVPEKPEIEKVDQPKI 725
OY 376 VSQVRAKYGEGYF-----EEKGISRYVFAKDLSEYTK--NLESKLSQESVSHLTAK 428
D 726 VD--LQIIIEPNFVDSQPNPQEPVEPSFVKIEKVPSEENKHASVDEVEKEKEVSEVEEK 783
OY 429 K-----ENVAPRDOEFYDKAVNLTBAHKA-----LFXNKGKNSDFQALD 468
D 784 QNSQSVSEIIPNEDEFEEDVTEQDLDKTYDPRIVEVEIIPSELHENEVAHPRIVEIE 843
OY 469 KLERLNDESTNKEKLVLDLLAFLA-PITHEPRLKPSQIETDE-VRIAQDLAKYTT 526
D 844 EVFPEPNQNNFQEIINEDDKSAHQHEIVEVEIIPEDDKNEKEHEIVEVEIIPEDKN 903
OY 527 SDGYTFDEHDIIISDGDAYVTPHMGHSWIGKDSIDKEXAAQVYTEKGLIPSPDAD 586
D 904 EKG--OHEIVEVE--EILPE-----DDKNEKEHEIVEVEIIP--EDKN 942
OY 587 VKANPTGSAAIYVRVKGKRIPLVRLPYWEH-TVEYKNGNLIIPKHQHNKIFAMF 645
D 943 EKGQHEIVEVEIIPEDDKNEK-----VEHEIVEVE--ILPEDKNEKG----- 983
OY 646 DDHTYKABNGYTLLEDFATIKYVEHPDE-RPHSNDGNGNAS---EHVLGKQHS- 697
D 984 -OHEIVEVEIIPEDDKNEKQHEIVEVEIIPEDDKNEKQHEIVEVEIIPEDDKNEKQ 1042
OY 698 -----DKNKFKADEPVEETPAEPVQVETEKVEAQKAEAVLLAKYDSSLK 747
D 1043 HEIVEVEEILPEDKNEKQHEIVE--VEEILPEDKNEKQHEIVEVEIIP- 1095
OY 748 ANATETLAGLRNNLTLOIMDNNSIMAAEAKLLALKGSNPSVSREKI 795
D 1096 -IEEVPQNTNNENIE---TIPBEKK-----NERSVEKAI 1128

RESULT 34
O9GTXT2 PRELIMINARY; PRT; 1236 AA.
AC O9GTXT2.
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Glutamate-rich protein.
GN
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_TaxID=5833;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FCI/NN.
RA Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M.;
RT "Molecular cloning and structure of the glutamate rich protein (GLURP)
RT gene of Plasmodium falciparum isolate FCI/NN";
RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.

```

```

DR EMBL; AF247634; AAG12326.1;
DR InterPro; IPR005479; Case L D2.
DR PROSITE; PS00867; CPSASE 2; 1.
SQ SEQUENCE 1236 AA; 141462 MW; 6438700909D34624 CRC64;

Query Match 4.2%; Score 175.5; DB 5; Length 1236;
Best Local Similarity 17.8%; Pred. No. 0.16;
Matches 155; Conservative 146; Mismatches 343; Indels 225; Gaps 37;

OY 7 YQATVVENRNVSTIDKQATOKTENLTP-----DEVSRREGINAEQVIKTIIDGCV----- 58
D 373 HNINVLDENNNNHNPQEPQEKPIESFEKPIDSEIILPEVETEEIIDDVSPGHSNHE 432
OY 59 ----VTSCHDVHYNGKVPYDAIISEELMKDQNYKLKEDIYNEVKGVYIKVDGKY 114
D 433 TFEETSESEHEAVSEKNAHEIYHEETVQGESNPEKADND--GNNSQNNNELNENEF 490
OY 115 VYLKDAAH--ADVNRTEEINROKQHSQ-----HREGTFR---NDGAVALLARSQGR 162
D 491 VESKSEHEPAEN-----EESLSEGHHEIIVPEQNNSESGSKLVNDGCGFEEAHHENF 546
OY 163 YTTDDGYTFNASDIIEDTGDAYIVPHGDHYTI-----PKNELASSELAAEAFLSG 214
D 547 SSEVSNSELENNEVE--SDKSVTEPAHEVEVSESNPEPAENESSSIEEAOEIVP- 603
OY 215 RGNLSNRTYVRQNSDNTSRTNNVPSVSNPCTTNTNTSNNSTNSQASQSDNDISLLKOL 274
D 604 -----EQNDESGES-----GLVNEBEGDFEERPHNEEPDPNDSELS- 642
OY 275 YKPLSGRHHVESDGLVFPDPAQITSRTARGVAVPHGDHYHPIYSQMSLEBRIARIIPLR 334
D 643 -----NELVSEKSVSEPAHEVEIIVSEKASBPAAHEVEIIVSEKSVSEPAHV----- 689
OY 335 YRSNNHWVPSRPEQSPQPTPEPSPGQAPANL---KIDSNSLSQVYAKVGEYGF- 389
D 690 -----ESVSEQSNNEPS-EKKDGPVSPFEEIEKVDVQPKIVD--LQIIIEPNFVDS 738
OY 390 ----EEKGISRYVFAKDLSEYTK--NLESKLSQESVSHLTAKK-----ENVAPRDO 437
D 739 QPNQEPVEPSFVKIEKVPSEENKHASVDPVEKKEVSEVVEKQNSQESVEEIPVNE 798
OY 438 EFDYKAVNLTBAHKA-----LFXNKGKNSDFQALD 467
D 799 EFDVHTFEQDLDDKTYDPRIVEVEIIPSELHENEVAHPRIVEIIEVPEPNQNNFQEI 858
OY 468 DK-----LRLNDESTNKEKL-----VDDLAFAPITHEPRLKPSQIET 509
D 859 NEDDKSAHQHEIVEVEIIPEDDKNEKHEIVEVEIIP-----PE--DKNEKQV 907
OY 510 YTEDEVRIAQDLADKYTTSDGYTFDEHDIIISDGDAYVTPHMGHSWIGKDSIDKEXAA 569
D 908 HEIVEVEEILPEDKNEKQHEIVEVEIIPEDDKNEKQH-----EIVEVEIIP 957
OY 570 QAYTKEG-----ILPPSPDADYKAN-----PTGDSAAAIYVRVKGKRIPLV 612
D 958 EDDKNEKQHEIVEVEIIPEDDKNEKQHEIVEVEIIPEDDKNEKQHEIVEVEIIPED 1017
OY 613 RLPYWEH-TVEYKNGNLIIPKHQHNKIFAMDDHTYKAPNGYT--LEDFATIKYVE 670
D 1018 DKNKFKADEPVEETPAEPVQVETEKVEAQKAEAVLLAKYDSSLK 747
OY 671 HPDERPHSNDGNGASHEVLGKQHSDEPNKFKADEPVEETPAEPVQV-ETEKV-- 727
D 1063 VPSQT-----NNENIEITIK--PEEKKNFESVBE--KAIQEPVPTLNENENVTP 1110
OY 728 ---EAOLKEAEVLLAKYDSSILKANATET 753
D 1111 KPSEGESTKPIVQIKVQEN-KENKKET 1138

RESULT 35
O8IHNS PRELIMINARY; PRT; 5507 AA.
ID O8IHNS

```

AC 081H3;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Antigen 332, putative.
 GN PF11_0507.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NCBI_TaxID=36329;
 RN [1]
 RC SEQUENCE FROM N.A.
 RC STRAIN=3D7;
 RX MEDLINE=2255705; PubMed=12368864;
 RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyte S.,
 RA Chan M.-S., Nene V., Shallow S.J., Suh B., Peterson J., Angiuoli S.,
 RA Perceva M., Allen J., Selengut J., Haft D., Mather W.W., Valdivia A.B.,
 RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,
 RA Fraser C.M., Barrett B.;
 RA "Genome sequence of the human malaria parasite Plasmodium
 RT falciparum.";
 RL Nature 419:498-511 (2002).
 DR EMBL; AE014843; AAN36076.1; -;
 SQ SEQUENCE 5507 AA; 621156 MW; 03B095165D1490AE CRC64;

Query Match 4.2%; Score 175.5; DB 5; Length 5507;
 Best Local Similarity 18.9%; Pred. No. 1.3;
 Matches 167; Conservative 163; Mismatches 337; Indels 215; Gaps 41;

QY 5 GYQARTVENNRVS---YIDGKQATQ---KTEMLTPDEVSKREGINAQIVIKITDQGY 58
 DB 2564 GGFTEIDVEEBSVTEIYVDESVTKIYEDDELTEIYVEBEGSFTEITV---EDBS 2620
 QY 59 VTSHGHHYHYNGKVPYDAIISEELMKDPNYLKDIEDYNEVKGQYIK--VDGKYY-- 114
 DB 2621 FT-----EEVIERSLIE--VEDTEYAEKESSSVIKETIDKSLTE 2661
 QY 115 -VYLDAAADNVRTYEINRQKQEHQSHREGTTPNDGAVALAROGRYTT---DDGY 169
 DB 2662 KIVEEKSYTEEVEEKSVEEVEBQRLVVEEBSATGIV---EDRLATEGIYDD-- 2714
 QY 170 INASDIID--TGDYIYPHGDHYHYIKNELSASEL-----AAEAFLSGKGLNS 221
 DB 2715 ILVTEIYEDGLATDFEVEQGGIIEVLDDEGSVTEIYVEEBSNEEIVEGVYI--- 2771
 QY 222 RYRQNSDNTSRNTWVPSVNPQTNTNTSNNSTNSQASQNDIDSLKQLYKL--PL 279
 DB 2772 ----BEDN-----IEPSEIIVGSVTEEMIKGLENEVILDDSDITEALEKESG 2820
 QY 280 SQRHVESDGLVPDPAQITSRTAGVAVPHGDHYHT--PYSONSELERIARIIPRYRS 337
 DB 2821 SEEIYEMGSLTEIYDEERSTSEDMIEGSSASEEIIQOESQVEEVEEVSVDIEI-- 2876
 QY 338 NMHWPSRREQSPQPTPEPSPEQOPAPNLKIDSNSIVSQVARKGEGVPEEKISRY 397
 DB 2877 ----VEEDLDKVEVEIEEFTVEVEHK--EEGSAVEIEVOEKEGSAVEE-- 2924
 QY 398 VFAKDLPEETVKULESKSKQESV---SHLLTAKKENV-----APRQGFYDKAVN 445
 DB 2984 VEEQAEVEEYVINEELIKESSDVEDVAKENELMNEEVEETQSAENNEEDKEIDNYVEE 3043
 QY 477 -ESTNEXKLVDDLAFLAPITHERKQKPSQIETEYDEVRINQDLAKTTSQGYFDDH 535
 DB 3044 TSEVTVEEVADEV-----PNSK--EVEIEESTIEIYEDGILTDLDVGQGG 3087
 QY 536 DIISDEGDAYVTTPHMGSHWIGKDSLDEKXVAAQAYTEKGIPLPSPDAD--VKANPTG 593

DB 3088 SYIEE-----VVEEVG-----SDSEEIYEBASITEVEKESVTEJDLVEESYTG 3132
 QY 594 D--SAAAIYNRKGEKRIPVLVLPVMEHTVEVKNGLNL--IPHGDHYINIKFAMWDDHT 649
 DB 3133 DILVSGSVTEEVVGBEK-----LVSEIYTEEGSAQEIYED-- 3170
 QY 650 YAPANGYTLLEDLFAITIKYVEHDEPRPSND---QWGNASEVHLKQHSDDPNKPKFAD 706
 DB 3171 --APATEEIDELIESTEEVVE--EBCPVDEIYVQEGSVTEETIIGESVEEVEEQQSGE 3226
 QY 707 EEP--VEETPAPEPEVQVE--TEKVEAQLKEAEVLL--AKYTDSSLKAN--ATEETLAGIR 758
 DB 3227 NEIEFVEEVSASQEIYQNESGTEETILEKVSASQEIYQDSVTEQIIEBQKPVTEEVNNE 3286
 QY 759 NNLTIQIMDNNSIMAEKTLALLGS-----NPSYSKE 793
 DB 3287 ESTIHEIIQES--HVEKVVQ--QGSVAEEVEPEVPSVTEE 3323

RESULT 36

ID 081CV5 PRELIMINARY; PRT; 1063 AA.

AC 081CV5;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Hypothetical protein.
 GN MAL6P1.63.
 OS Plasmodium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 NCBI_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Cherevach I., Davis P., Goodhead I., Stevens K., Mungall K.,
 RA Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
 RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrett B.;
 RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AL844505; CAD50334.1; -;
 KW Hypothetical protein.
 SQ SEQUENCE 1063 AA; 123332 MW; 2A2677237A74C755 CRC64;

Query Match 4.2%; Score 173; DB 5; Length 1063;
 Best Local Similarity 18.2%; Pred. No. 0.18;
 Matches 182; Conservative 157; Mismatches 361; Indels 298; Gaps 45;

QY 14 ENNRVSYIDGKQATQK--TENLTPDEVSKREGINAQIVIKITDQGYVT-----S 61
 DB 45 ENSLNFCKGVSNSNEELSYDLSNRRMSDISGLRKNNGIKVNVNLLITQSRINNINENT 104
 QY 62 HGDHYHYNGKVPYDAIISEEL--LMKDPNYLKDEED-- 97
 DB 105 NCNVNNKNKKEELNVNMINNINDEYKESNFIILDEDEYIKNDTYISCTDVNNESEKTR 164
 QY 98 -----VNEVKGQYIVKVDGKYVYLKDAHADNVYTK--EINRQKQEHQSHREG 146
 DB 165 NSDNNFNNSYITSSANNNVVSKSGSY---KCPAENNTMTMKRLSDINNINNINNINN 217
 QY 147 TPRNDGAVALARSQGRYTTDDGYIFNAS-----DILBDTGDAYIVPHGDHYHY-- 194
 DB 218 --NNDNNINIGNEGK--NNYFLFNNNKDGRLRRSSLFNLKSKSCNFFLPDDGFYNNAN 272
 QY 195 -----IPKNELSASELAAEAFISGKGLNSNR---TYRQ 227
 DB 273 NNNNNNNNNNNKNSNYISNDFSKKLLINKKNDHTFL-----LSNKKIYDYTKSR 327
 QY 228 NSD-----NTSRTWVPSVNPQTNTN-----TSNN--SNTNSQASQNDIDSL 270
 DB 328 KSDKKEFITNTDNNSEGSVIFPEMEKRYNFTLPSEKITSYVPLKKNSTPPSLKEDML 387
 QY 271 LKQLYKL-----PLSQRHVESDGLVPDPAQITSRTAGVAVPHGDHYHFIYS 318
 DB 388 LKSKYSINNNKNN 444

QY	3139	QMSLEEBERARIIIPLRYSNMHWPPSRPQSPQTPPESPQPGAPAPNLIKDSNSLVS-	377
DB	445	LSGEHNYKIGSWSMETYNRNTDV-DKSDSEPLSE-----QNEIDNHVLLNR	491
QY	378	-----OLVAKVOEGY--VFEKGISRYVFAK-DLPSEIVKNLESKLKQESVSHTL--TA	427
DB	492	SKLLIEISKKCLSGVSDVASNMNMDLIYKKKGRNNSTYDNI CDSPFKQGESVSKLSLPSD	551
QY	428	KKENVAAPRD-----OEFDKA-----YNLITE--	449
DB	552	KOKQIEPIEISKDNDNNNNNNKNSIDOLLNLKMTDSNGDYVNHMDNVQSYSDTEED	611
QY	450	-----AHKALFXNKGNSDPOALD-----KLLERLNDSTYKKEKLVDDLAFAP1	495
DB	612	EDDNDNDNCRGQKSKSIDHGLMNDVNEQNI CNLKNVATNQNNNNNNN--DNISKFKE--	667
QY	496	THPELGRNSQIEYTEDEVRIAQ---LADKY-----TTSDGY-----	530
DB	668	EHIEKNGKKINTIPEAYDKNITOGKEMAYKVNINNYMNTGGDNNITGVNNSGHEKKKE	727
QY	531	--TFPEHDIIISDEGAIVYTPHKGHSHWICKDSLSDEK-----YAAQATYKE--KGI	578
DB	728	IKIEVCSISKSEEDALI--NIDRNHVLLDHPKKEEKVESGGGANNMIYTKESHDKNI	785
QY	579	LPPSPDADKAPPTGDSAAAIYNRKGEGRIPVLBPVVEHTVEVKGKGNLIIPHKDHYN	638
DB	786	VSDDHNIID-----DDDDDIYIEKLDKSHL-----TEKDPVAVADVRYN-----DEKE	831
QY	639	NIKFAWFDHTYKAPNGYLTLEDLFTIKYYVHPD--ERPHSNDGWNASEVHLKKDHS	696
DB	832	SI-----SLDGTGSARKIKIHPDPFEVEKYLDN---KKNDVSMGYREBI	871
QY	697	EDPNKNGFADPEPVEETAPEPVPQVEITEKVAQ--LKEAEVLAKYDSSILKAN-AIET	753
DB	872	KISDQKALDKSVNVINYSNDVKLFREKYFLQDLCEKRWIKWIOKSKIENTKNTIIEIKS	931
QY	754	LAGLRNNLTLOIMDNNS-----IMAEAEKIL-LALLKGSN	786
DB	932	LISFNNNISYALKEKDAEIKIKKKQKELEIYIMKRVN	969
RESULT 37			
QY	Q9U459	PRELIMINARY:	PRT: 5458 AA.
AC	Q9U459	O8GPG4:	
DT	01-MAY-2000	(TEMBLrel. 13, Created)	
DT	01-JUN-2001	(TEMBLrel. 17, Last sequence update)	
DT	01-MAR-2003	(TEMBLrel. 23, Last annotation update)	
DE	Erythrocyte membrane-associated giant protein antigen 332.		
OS	Plasmodium falciparum.		
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.		
OX	NCBI_taxid=5833;		
RN	(1)		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=FCFC/NH;		
RA	Shan Z.X., Yu X.B., Li X.R., Ma C.L., Fang J.M., Lu J.H., X J.;		
RT	"Molecular cloning and structure analysis of the plasmodium falciparum		
RT	erythrocyte membrane-associated giant protein Ag332 (Pf332) gene.";		
RL	Submitted (Feb-2001) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF202180; AAF15293.3; -		
DR	InterPro; IPR006763; AG332.		
DR	InterPro; IPR001313; Pumlilo/Puf.		
DR	Pfam; PF04671; AG332; 138.		
SO	SEQUENCE	5458 AA; 615269 MW; 4F6A37D92BD80172 CRC64;	
Query Match 4.1%; Score 172.5; DB 5; Length 5458;			
Best Local Similarity 18.9%; Pred. No.1.9;			
Matches 167; Conservative 162; Mismatches 338; Indels 215; Gaps 41.			
QY	5	GLYQARTYKNNRVS---YIDSKATQ--KTENLTDPDEVSFKREGINAEQIYIKTTDQY	58
DB	2629	GSFTDVIIEEBSVIEEIVVEBSYTKIEVEDEELVTEIIVDESGFTPEIY---EDGS	2685

QY 59 VTSHGDNHHYNGKVPYALIISEELIMKDPRYKDKDEIVNEVGVYIK--VQKYY-- 114

Db 2666 FT-----EEVIEEESLLEE-----VEDTETVAKEBGSVYIKETIDEKSLTE 2728

QY 115 -VYLKDAHAADNAVTKEEINRQKQSHQSHRECGTTPRDGVALARSGQRYTT---DDGY 169

Db 2727 KIVEEKSVTVEEVEKESVKKEVEEQRUVVEEESATGIV-----EDRLATEGIVDD-- 2779

QY 170 IFNASDIED--TGDATVYPHGDHYHYPKVELSASEL-----AAAPAFISGRGNLSNS 221

Db 2780 ILVTEIEIVEDGLATDFEVOQSGIIEEVLDEGSGVTEIEVEEGSPMEIVEGVSVI--- 2836

QY 222 RTYRQRNSDNRSTNMPVSVSNPGTNTNTNNSNTNSQASQNSDIDSLTQLYKL--PL 279

Db 2837 -----EEEDN-----IIPVSEIVEEGSVTEEMIKEGLENEVIDDESI TEBALEKESV 2885

QY 280 SQRHVESDGLVDPRAQITSRTAGVAAPHGDHYHFI--PYSQMSLEERARIIPRYRS 337

Db 2886 SEEIVIEEWGSLTEIEVIDERSTSDMIIEGSGASEEIIIEGSGVEEVEEVSVIDEI--- 2944

QY 338 NHMVPDSRPEDPSQPPTPEPSPGQAPRNLIKIDNSSIVSLQYARKVGGYVPEEKGISRY 397

Db 2942 -----VEDELDITKEVBEIEIFNTEVEVEHK--EEGSAVEEIIVOEEKGSVNER-- 2989

QY 398 VFAKDLPEYTKNLESKSKQESV---SHTLTAKENV-----APRDOEFYDKAYN 445

Db 2990 -IIEVGSITTEMEBODVSDNEIEVEERSVIEEAEENWMIKEVBEESGDNEEVIYDEDS 3048

QY 446 LLTEAHKALFYXK---GRNSDFOALDKLERLND----- 476

Db 3049 VSEQAEVEVYINEELIKESSDVEDVUKVENELMNEVEEQTQSVANNEDKELDNYVEE 3108

QY 477 -ESTNKELVDDLLAFLAPITHPERLGRPNQIETDEEVRVIAQLADKYTTSDGYIPFH 535

Db 3109 TESVTEEVVNDV-----PNSK-EVQETESIIIEEIVIEGLTITDLVGOQS 3152

QY 536 DIISDEGAYVTPHMGSHWIGKDSLDPKEVAAQAATKEGILPPSPDAD--VKANPTG 593

Db 3153 SVIEE-----VVEEVG-----SDSEIVEEASTTEVEEKKESSTEDIIIVESVTG 3197

QY 594 D-SAAALYNKVGEBKRIPLRVLPYMEHTVEVKNGLI--IPKHDAHNIKFAWEDDT 649

Db 3198 DILVEGSTEVEEVEEK-----LVSEIIVTEEGSVAQEIIVED----- 3235

QY 650 YKANGNYLTEDLFATIKYVVEHPDERPHNSD--CWGNAASHVLGKKDHSBDPMKPFKAD 706

Db 3236 --APATEIEDIESVTEEVAE--EEGPDEEIVVOEGSVTEIEIIQGESKVEEVEEEOGSE 3291

QY 707 EEP--VEETPAPEPEVQVE--TEKVEAQILKAAYLL--AKYTDLSLKAN--ATETLACLAR 758

Db 3292 NEELFVEEVSASQELIVQNESGTEIEILKRVASQELIVDQGSTEQIIEQKRVTEEVAE 3352

QY 759 NNLTLOIMDNNISMAEAKLIALKGS-----NPSSVSKKE 793

Db 3352 ESITHETIIQES--HVEKVVQ--QGSVAEEVENPVSVTEE 3388

RESULT 38

Q9LJC9 PRELIMINARY; PRT; 1795 AA.

AC Q9LJC9; 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, last annotation update)

DE Truncated fmbB.

GN FMTB OR TRUNCATED FMTB OR MW2087.

OS Staphylococcus aureus, and

OC Staphylococcus aureus (strain MW2).

OC Bacteria: Firmicutes: Bacillales; Staphylococcus.

NCBI_TaxID=1280, 196620;

RP SEQUENCE FROM N.A.

RC STRAIN=KSA8;
 RA MEDLINE=20348625; PubMed=10896508;
 RX Komatsuzawa H., Ohta K., Sugai M., Fujiwara T., Glanzmann P.,
 RA Berger-Bachi B., Suginaka H.;
 RT "tnb51-mediated insertional inactivation of the fmbB gene encoding a
 cell wall-associated protein abolishes methicillin resistance in
 Staphylococcus aureus.";
 RL J. Antimicrob. Chemother. 45:421-431(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MM2;
 RX MEDLINE=22040717; PubMed=12044378;
 RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
 RA Nagai Y., Iwano N., Asano K., Naito T., Kuroda H., Cui L.,
 RA Yamamoto K., Hiramoto K.;
 RT "Genome and virulence determinants of high virulence community-
 acquired MRSA.";
 RL Lancet 359:1819-1827(2002).
 DR EMBL; AB015223; BAA93430.1; -;
 DR EMBL; AP004829; BAB95952.1; -;
 DR InterPro: IPR000890; Acetate kin.
 DR InterPro: IPR005877; Gpos_YSTRK.
 DR Pfam; PF04650; YSTRK_signal; 1.
 DR TIGRFAMs; TIGR01168; YSTRK_signal; 1.
 DR PROSITE; PS01075; ACETATE_KINASE_1; 1.
 DR Complete proteome.
 SQ SEQUENCE 1795 AA; 191008 MW; 6CB77CCFCB3D350 CRC64;

Query Match 4.1%; Score 172; DB 16; Length 1795;
 Best Local Similarity 19.5%; Pred. No. 0.42;
 Matches 169; Conservative 142; Mismatches 355; Indels 202; Gaps 38;

QY 8 QARTYKNNRVSYIDGKATOKT-----ENLTPDEVSRREGINAQIVIKITDGYVNS 61
 DB 989 QATTKKSDAKAEIKQKSEKRTAIEAMNDSTTEBQAKKXV-----DQAVLTA 1038
 QY 62 HGD-----HYHYNGKVPYDAIISEELMKDPNYLKDEDIVNEVKGGVYIKVDGKYV 115
 DB 1039 NADINDAANNVDNAKTNTNEATIA--AITPDANVY-----PQAKQAIADVQAOQ--- 1086
 QY 116 YKDAHADNVTKKEINRQKQEHSHRGCGTFRNDGA-----VALARS-----OG 161
 DB 1087 --EKAIIDANNGSTTEKAAKQOVQTEKTTADAIDAATNAEVEAAKAEIAKIEAIOP 1144
 QY 162 RYTTDDGYIFNASDIIEDTGDAIYVPHGHHYIPKNEISASSELAAAEFLSGRLNS 221
 DB 1145 ATTTDQ---NAKEAIAATKANERKTA-----IAQTQDITAEIAAANA-----DVDNA 1188
 QY 222 RTYRRQNSDNTSRITNVVPSVNPCTNTNTNSNNTNSQASQNSDIDSLIKOLYKLPLO 281
 DB 1189 VTQANSNINAAANSQNDVDQAKTTGERSID-QVTPYVKKATARNETITALINN--KLQETIO 1245
 QY 282 RHVESDGLVFPDPAQITSRFAGVAVPHGHHYFIPYSQSELEERARIIPRYRSNHV 341
 DB 1246 ATPDARDEKQADAAEANTENGKAI-----QAIATATNAQVDEAKTNAEAAI 1293
 QY 342 PPSRPEQSPQPPPESSPQPORA-----PYUKIDSNSLSVQVRYKVGES-----YV 388
 DB 1294 NAVTPPVVKKQAKBEIDQLQATQTNVINNDQATTEKEEAAIQOALATVTTAKNNITAA 1353
 QY 389 PEEKISRYVFAKDLPSETVKNLES-----KLSKQESVSHTLTPAKKENVAPRDEPFYKAY 444
 DB 1354 TDDNGVDQ---AKDAGKNSIQSTQPAITSYKSNAKNDVDQAVTTQNOAI-----DNTT 1402
 QY 445 NILTEAHKALFYNKGNSDFQALDKLLER-LNDESTN-----KEKLVDDLAFLEATIPH 498
 DB 1403 GATTEE-----KPAADLVLKAEKAYODILNAQTNDVTQIKDQAVADIGITADPTIK 1457
 QY 499 EHLGKNSQIEYTEDEVRIAQALADKTTSDGYIFDEHDIISDEGDAVYVPRHMGHSWIGK 558
 DB 1458 D--VADELATKANEQOALIAQTDART-----EKEQANOQVDAQLT--QGNQNIENA 1507
 QY 559 DLSLDEKVAQAQAYTEKGIILPPSPDADVANKPTGDSAAAIYVRVGEKRIPLVRLPVNV 618

DB 1508 QSIDDVNTAKDNAL---QAIIDPIQASTDVKTARAEILTEMQNKI-----TEILNAN 1556
 QY 619 EHTTEVKNQNLIP-----HKDHYNIKPAMF--DHTYKAPNGYLTEDLFAITIKYVEIP 672
 DB 1557 ETTTNEEK-GNDIGPRAVAYEEGLNNINAAITTGDTTADTAVQKVOQLHA-----NP 1608
 QY 673 DERPSNDQGNASAEHVQCKDHSEDPNNKFADEE-PVEETP--AEPEVPQVETEKVEA 729
 DB 1609 VKKP-----AKKELDQ-----AAADKQTQIEQIPNNSQDEINDAKOE-VDT 1649
 QY 730 QKEA-----EVLAKVYIDSLSKANATETLAGLRNNLTLOIMD----- 767
 DB 1650 ELNQAKTNVDQSTNEVYDNAVKEGKAKINAVKTFSEYKDALAKIEDAYNAKNVEADNS 1709
 QY 768 ---NNSINAEAKLALALKGSNPSYSK 792
 DB 1710 NASTSSEIAEAKOKLAEIKOTADQVNVQ 1737

RESULT 39

ID Q9LCH2 PRELIMINARY; PRT; 2478 AA.

AC Q9LCH2;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE FmbB.

GN FmbB.

OC Staphylococcus aureus.

OX Bacteria; Firmicutes; Bacillales; Staphylococcus.

NCBI_Taxid=1280;

RP SEQUENCE FROM N.A.

RC STRAIN=COL;

RX MEDLINE=20348625; PubMed=10896508;

RA Komatsuzawa H., Ohta K., Sugai M., Fujiwara T., Glanzmann P.,

RA Berger-Bachi B., Suginaka H.;

RT "tnb51-mediated insertional inactivation of the fmbB gene encoding a

cell wall-associated protein abolishes methicillin resistance in

Staphylococcus aureus.";

RL J. Antimicrob. Chemother. 45:421-431(2000).

DR EMBL; AB025716; BAA93438.1; -;

DR InterPro: IPR005877; Gpos_YSTRK.

DR InterPro: IPR001899; Gram_pos_anchor.

DR Pfam; PF04650; YSTRK_signal; 1.

DR TIGRFAMs; TIGR01167; LpYTG_anchor; 1.

DR TIGRFAMs; TIGR01168; YSTRK_signal; 1.

SQ SEQUENCE 2478 AA; 262995 MW; 1C118EBE0DB03B34 CRC64;

Query Match 4.1%; Score 172; DB 2; Length 2478;
 Best Local Similarity 19.4%; Pred. No. 0.66;
 Matches 169; Conservative 141; Mismatches 354; Indels 206; Gaps 38;

QY 8 QARTYKNNRVSYIDGKATOKT-----ENLTPDEVSRREGINAQIVIKITDGYVNS 61
 DB 986 QATTKKSDAKAEIKQKSEKRTAIEAMNDSTTEBQAKKXV-----DQAVLTA 1035
 QY 62 HGD-----HYHYNGKVPYDAIISEELMKDPNYLKDEDIVNEVKGGVYIKVDGKYV 115
 DB 1036 NADINDAANNVDNAKTNTNEATIA--AITPDANVY-----PQAKQAIADVQAOQETAIOG 1089
 QY 116 YKDAHADNVTKKEINRQKQEHSHRGCGTFRNDGA-----VALAR-----SGG 161
 DB 1090 -----NNGSTTEKAAKQOVQTEKTTADAIDAATNAEVEAAKKAIAKIEAIOP 1141
 QY 162 RYTTDDGYIFNASDIIEDTGDAIYVPHGHHYIPKNEISASSELAAAEFLSGRLNS 221
 DB 1142 ATTTDQ---NAKEAIAATKANERKTA-----IAQTQDITAEIAAANA-----DVDNA 1185
 QY 222 RTYRRQNSDNTSRITNVVPSVNPCTNTNTNS--NSNTNSQASQNSDIDSLIKOLYKLPLO 279
 DB 1186 VT---QANSNINAAANSQNDVDQAKTTGERSIDQVTPYVKKATARNETITALINN--KLQETIO 1240

```

OY 280 SQRHVESDGLVDPDPAQTSRTARGAVPHGDHYHIFPYQMSLEBRIARIIPLYRGNH 339
D 1241 IQATDADAEDEKQADAEANTENGKA-----NQASATTAQVDEAKANEA 1288
OY 340 WVPDSRPEQSPQPTPEPSPGPOPA-----PMLKIDSNSLSVQLVRKYGEG- 386
D 1289 AINAVTTPKVVKKQAQKDEIDQATQTVINNDQATTEKEKAALQOLATAVTADAKNNIT 1348
OY 387 YVFEKKGISRYVAFADLPSEYTKNLES-----KLKSQESVSHLTAKKENVAPRDEFEYDK 442
D 1349 AATDNGVDQ---AADAGKNSIQSTQPAATVAVKSNKNDVDAQVTTQNOAI-----DN 1397
OY 443 AYNNLTTEAHKALFXKKGNSDFQALDKLER-LNDESTN-----KEKLVDDLAFAPIT 496
D 1398 TTGATTEB-----KNAKDLVLAKAKEKAYODILNQTTNDVTOIKDQAVADIQGITADTT 1452
OY 497 HPERLGKPNQSEYTEDEVRIAQLADKYTSDGYIFDEHDIISDEGDAVYTPHMGSHMI 556
D 1453 IKD-VAKDELATFRANEQKALIAQTADATT-----EKEQANQGVDAHLT--QGNQNI 1502
OY 557 GKDISLDEKVAQAQVTEKEGILPPSPDADVYKANTGSAALIVRVKGEKRIPLVRLPY 616
D 1503 NAQSIDDVNTAKDNAI-----QALDPIQASTDVYKTNARABELTEMONKI-----TEJLN 1551
OY 617 MVEHTVEVKNGLIIP-----HKDHYNIKFAMF--DDHTYKAPNGVYTEDLFATIKYVYE 670
D 1552 NNETTNEEK-GNDIGPVRAAYEGGLNNINAAITTTGDDVTAKDTAVOKVOQLHA----- 1603
OY 671 HPDERPHSNDGWSNASEHVLGKDHSEDPNNKFKADEE-PVEETP--AEPEVQVETEKV 727
D 1604 NPVKRP-----AGKKELDQ-----AAADKKTQIQIEQTNPASQOEINDAKQ-E 1644
OY 728 EAQLKEA-----EVLAKVYTDSSILKANATETLAGLRNNLTQIMD----- 767
D 1645 DTELNAQTATNVDOSSSTNEVDNAVKEGAKINAIVTFSEYKKDALAKIEDAYNAKVNEAD 1704
OY 768 -----NNSIMAEKLLALLKGSNPSSYSK 792
D 1705 NSMASTSEIAPKQKLAELKOTADONVNO 1734

RESULT 40
O99GR6 PRELIMINARY: PRT: 2481 AA.
AC O99GR6 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE 3meB protein.
GN 3meB (MRP) OR FMTB OR SAV2160 OR SA1964.
OS Staphylococcus aureus (strain Mu50 / ATCC 700699), and
OC Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878, 158879;
RN [1]
RP SEQUENCE FROM N. A.
RC SPECIES=S. aureus (strain Mu50), and S. aureus (strain N315);
RX MEDLINE=21311952; PubMed=11418146;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
RA Mizutani-Uji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaico C.,
RA Sekimizu K., Hiraoka H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hattori M., Ogasawara N., Hayashi K., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
RT aureus.";
RL Lancet 357:1225-1240(2001).
DR EMBL, AP003364; BAB58322.1; -.
DR EMBL, AP003136; BAB43253.1; -.
DR InterPro, IPR000890; Acetate_kin.
DR InterPro, IPR005877; Gpos_Y5IRK.

```

```

DR InterPro, IPR001899; Gram_pos_anchor.
DR Pfam, PF04650; Y5IRK_signal; 1.
DR TIGRfam, TIGR01167; LPTXG_anchor; 1.
DR TIGRfam, TIGR01168; Y5IRK_signal; 1.
DR PROSITE, PS01075; ACETATE_KINASE_1; 1.
KW Complete proteome.
SQ SEQUENCE 2481 AA; 263765 MW; E1EAA899B81665E8 CRC64;

Query Match 4.1%; Score 172; DB 16; Length 2481;
Best local similarity 19.4%; Pred. No. 0.66; Indels 210; Gaps 38;
Matches 169; Conservative 140; Mismatches 333;

OY 8 QARTVKENNRVSYIDGQATQKT-----ENLTPDEVSKREGINAEQIVIKITDQGYVTS 61
D 969 QAAATTKSDAKAEIAQKASERKTAIEAMNBSSTEEQQAQKV-----DQAVTA 1018
OY 62 HGD-----HYHYNGKVPIDALISEELMKDPYKLDKEDIVNEVKGYIKYDKGYV 115
D 1019 NADIDNATANTDVNNAKTTEATTA--AIPDAVVKPAKQAIAD-----KVQAQ--- 1066
OY 116 YLKDAHADNVRTKEELNRQKQSHREGSTPRNDGAVLARS----- 159
D 1067 --ETALDANNSTTEBEKAQOVQTEKTA--DAIDAHSNVEVEAKNAEIAKIE 1120
OY 160 --QGRYTTDGYIFNASDIEDTDGDAYIVPHGDHYHYIPKNELSASEIAAAEAFLSGRN 217
D 1121 AIPATTTKD-----NAQALATKANERKTA-----IAQTQDITAEIIAANA-----D 1164
OY 218 LNSRTYTRQNSDNTSRTNWVPSVNPCTTNTTNSNNTSQAQSQSDIDSLKQLYKL 277
D 1165 VDNAVTQANSIEANSONDQAKTGETSID-QVTPVKKKATAREIETAILNN--KL 1221
OY 278 PLSQRHVESDGLVDPDPAQTSRTARGAVPHGDHYHIFPYQMSLEBRIARIIPLYRYS 337
D 1222 QETATPDATDEEKQAADAEANTENGKA-----NQASATTAQVDEAKANA 1269
OY 338 NHVWDSRPEQSPQPTPEPSPGPOPA-----PMLKIDSNSLSVQLVRKYGEG- 386
D 1270 EAINAVTTPKVVKKQAQKDEIDQATQTVINNDQATTEKEKAALQOLATAVTADAKNN 1329
OY 387 --YVFEKKGISRYVAFADLPSEYTKNLES-----KLKSQESVSHLTAKKENVAPRDEFEY 440
D 1330 ITAATDNGVDQ---AADAGKNSIQSTQPAATVAVKSNKNEVDQAVTTQNOAI----- 1378
OY 441 DKAVNNLTTEAHKALFXKKGNSDFQALDKLER-LNDESTN-----KEKLVDDLAFAP 494
D 1379 DNTTGATTEB-----KNAKDLVLAKAKEKAYODILNQTTNDVTOIKDQAVADIQGITAD 1433
OY 495 ITHPERLGKPNQSEYTEDEVRIAQLADKYTSDGYIFDEHDIISDEGDAVYTPHMGSH 554
D 1434 TTIND-VAKDELATFRANEQKALIAQTADATT-----EKEQANQGVDAHLT--QGNQNI 1483
OY 555 WIGDISLDEKVAQAQVTEKEGILPPSPDADVYKANTGSAALIVRVKGEKRIPLVRL 614
D 1484 IENAQSIDDVNTAKDNAI-----QALDPIQASTDVYKTNARABELTEMONKI-----TEI 1532
OY 615 PYWHEHTVEVKNGLIIP-----HKDHYNIKFAMF--DDHTYKAPNGVYTEDLFATIKY 668
D 1533 LNNETTNEEK-GNDIGPVRAAYEGGLNNINAAITTTGDDVTAKDTAVOKVOQLHA----- 1586
OY 669 VEPDERPHSNDGWSNASEHVLGKDHSEDPNNKFKADEE-PVEETP--AEPEVQVETE 725
D 1587 --NPVKRP-----AGKKELDQ-----AAADKKTQIQIEQTNPASQOEINDAKQ-E 1626
OY 726 KVEAQLKEA-----EVLAKVYTDSSILKANATETLAGLRNNLTQIMD----- 767
D 1627 -VDTELNAQTATNVDOSSSTNEVDNAVKEGAKINAIVTFSEYKKDALAKIEDAYNAKVNE 1685
OY 768 -----NNSIMAEKLLALLKGSNPSSYSK 792
D 1686 ADNSMASTSEIAPKQKLAELKOTADONVNO 1717

```



```

OY 8 QARTVKNRRVSYIDGKATQKT-----ENLTPEVSKREGINAEQIVIKITDGYVTS 61
DB 986 QAAATTKSDAKAEIINQKASERKATIAEAMNDSTTEEGQAAKDKV-----DQAVVTA 1035
OY 62 HGD-----HHYVGVKPYDAIIEBELMKDPNYKLDKEDIVNEVGKGYIVKDGKYYV 115
DB 1036 NADIDNMAAMNDYDPAKTTNETATIA--AITPDANKVPAKQAIADKVQAOETAIIDG----- 1089
OY 116 YLKDAAHADNVRTEKINRQKQESHQHRGGRPRNDGA-----VALAR---SGG 161
DB 1080 -----NGSTTEKAAAKQOVOTEKTADAIDAHAHTNAEVAKAAIAKAEIAP 1141
OY 162 RYTTDDGYIFNASDIIEDTGDAIYVPHGDHYHYPKNELASASELAEAENFLSGRNLNS 221
DB 1142 ATTKD-----NAKKAIAATKANERKTA-----IAQTODITAEIAAANA-----DVDNA 1185
OY 222 RTYRQNSDNTSRITWVSVSNPGTTNTYTSN--NSNTNSQASQNSNDLSLKQLYKPL 279
DB 1186 VT---QANSNIENANSQNDVDQAKTTGENSIDQVTPYVKKATAPANEITAILINN--KLQE 1240
OY 280 SORHVESGIVFDPQAQITSRTAGVAVPHGDHYHIFPVSQMSLEBRTARIIPLAVRSNH 339
DB 1241 IQATPDADDEKQADAEANTENGA-----NQAISAAITNAQVDEAKANABA 1288
OY 340 WVPDSRPEQSPQPTPEPSPPGPPA-----PULKIDSNSLSVQVVRKVGEG----- 386
DB 1289 AINATPRTVYKQAKADEIDQAOQVTVINNQNQAATTEKKAALQQLATATVTDKANNIT 1348
OY 387 YVFEKGISRVFAKOLPSETVKNLES---KLKQESVSHLTJAKKENAVAPDOEFYDK 442
DB 1349 AATDNGVDQ---AKDAGKNSIQSTQPAIVAKSNKNDVDQAVTTONQAI-----DN 1397
OY 443 AYNLTTEAHKALFXKKGNSDFQALDKLER--LNDESTN-----KEKLVDDLAFAPIT 496
DB 1398 TTGATTEE---KNAADVLVAKAKAVQDITLNQTTNDVTQIKDQAVADIQGITADTT 1452
OY 497 HPERLKQNSQIETEDEVRIQAQADKYTSDGYIFDEHDIISDEGDAVVTPRMGSHWI 556
DB 1453 IKD--VAKDELATKANEQKALIQATADAT-----EKEQANQGVADQLT--QGNQMB 1502
OY 557 GKDSISDEKVAQAAYTEKKGILPPSPADAVKANPTGSAALIVRVKGEKIPLVRLPY 616
DB 1503 NAOISIDVNTAKDANI---QAIDPIQASTDVKTNAARELLTETMOKKI-----TEILN 1551
OY 617 MVEHTEVXNGKLLIP-----HKDHNIKKAMF--DHTYKAPNGVTLEDLPAITIKYYVE 670
DB 1552 NNETNEER--GNDIGPVAAYEELNINAAITTTGDVTTAKOTAVQVOQLHA----- 1603
OY 671 HPDERPHSNDGWNASEHVLGKDKHSEDPNKNFKADEE--PVSETP--AEPEVPQVETEKV 727
DB 1604 NPVKP-----AGKKEIDQ-----AADKKTQIQETNNAQQOEINDAKQE--V 1644
OY 728 EAQKKA-----EVLAKVTDSLIKANAETTLGLRNNLTLOIMD----- 767
DB 1645 DTELQAQNTNVDQOSTNEVNDVAVKEGAKAIAAVTFSEYKKDALAKIEDAYNAKVNED 1704
OY 768 -----NSGMAEAKTLALIKGNSPSSVSK 792
DB 1705 NSMASTSEIAEKQKALQKOTADQNVNQ 1734

RESULT 43
O82345 PRELIMINARY; PRT; 1043 AA.
AC 082345;
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, last sequence update)
DT 01-MAR-2003 (TREMblrel. 23, last annotation update)
DE AT2G46240 protein.
GN AT2G46240.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

```

```

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
RA Cronin L.A., Shen M., Vanaken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nieman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
  thaliana";
RU Nature 402:761-768 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X.;
RU Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL, AC005397; AAC62882.1; -.
DR InterPro; IPR003103; BAG.
DR InterPro; IPR000048; IQ_region.
DR Pfam; PF02179; BAG; 1.
DR Pfam; PF00612; IQ; 1.
DR SMART; SM00264; BAG; 1.
DR SMART; SM00015; IQ; 1.
DR PROSITE; PS50096; IQ; 1.
SQ SEQUENCE 1043 AA; 116758 MW; B5E271A02299177B CRC64;

Query Match 4.0%; Score 168.5; DB 10; Length 1043;
Best local similarity 18.5%; Pred. No. 0.31;
Matches 169; Conservative 108; Mismatches 327; Indels 311; Gaps 38;

OY 40 KREGINAEQIVIKITDGYVTSCHGDHYHYNKXV---PYDAIIEBELMKDPNYKLDKED 96
DB 271 KRDPVEASE-----SSNEDRKMKQNGKTVVEYFPDITSMKSL----- 306
OY 97 IVNEVGKGYIVKDGKYYVILKDAHADNVRTKEI-----NRQKQESHQ 142
DB 307 ---IQ-----QDVEKQONQKKEPGQVPIPIWIPSGYGRKQVEASES 348
OY 143 REGGTP-----RNDGAVALARSGQRYTTDDGYIFNASDIIEDTGDAIYVPHGD 190
DB 349 KESNENGRNLESCSDLRHNGQITQAKGK-----EGNFECNVLSDAEKSSVINIPIVAN 403
OY 191 HY-----HYIPKNELASASELAEAFLSGR----- 215
DB 404 HLGEPRIPIVYKLSNHLPKPTPTKRIAKNPNVSKYKKEQSSSSSEASKLPVCLRVDP 463
OY 216 -----GNLSNRYRQNSNT-----SPNNWPSVSNPGTTNTNS--NN 254
DB 464 PKERNQSKSVSHRPRKEKSKETKIAAPLSKKAESRT--VPBACNVKCEBNAEMKMAE 521
OY 255 SNTNSQASQNSNDISLKQLYKPLPSORHVESDGLVFPDPAQITSRTAGVAVPHGDHYH 314
DB 522 GSIALAKRTKESVES-----NSNIOEESNBEIITPCAKXNRBP----- 561
OY 315 IYPSQMSLEERARIIPRYRS---NHVWDSRPEQSPQPTPEPSFG----- 360
DB 562 ---AKKSTEEBAARIISQMYRGYDVRWBIKYLKEIA---TVREGQGVKKRIEALEA 615
OY 361 ---PPAPNLIKDSNSLSVQVLRVVGEGYVEEKGISRYVAKOLPSETVKNLESKLS-- 416
DB 616 STDHIEKEKIVVNGEIVMNLTKLD--AVEGHPISIREFRKALATE--LSIDKIDSL 671
OY 417 KQESVSHLTJAKKENAVAPDOEFYDKAYNL---LTEAHKALFXKKGNSDFQALDKLE 472
DB 672 KNSCASAKEKAVKQVELKQSP--SDSPVNLHSQLTEENKRV----- 712
OY 473 RLNDSTNKEKLVDDLAFIAPITPHPRLGKPNQSIETEDEVRIQAQADKYTSDGYIF 532

```

Db	713	-----SDTULEKVR-----LSPEEH-----PMSVLANITDEK-----QMSALETBEQ--	751
OY	533	DEHDIISDEGAYVTPHMGSHWIGKDSLDEKUYAQAAYT-----KEKGILPPSPEDA	585
Db	752	-----GLFETLATDSKQATENNAASSTIPEKIGEVEVTVGNP--	791
OY	586	DVKANPTGDSAAAIYRKYGEKRI-----PLVRLPTMVEHT-----VEV--	624
Db	792	-----PSADGNGMTVTVAENKAMVSELEEBINELPOMVEETETNSIRDPENASVEVSEA	846
OY	625	-----KNGNLIIPHKD--HYNIKIFAMPDHT-----YKAPNGYTLJEDLFA-----	663
Db	847	ETNSSENENRKGEDDIVLHSEKANVELSELPGVVIDEFTQPLSDQBPSSYTRKGNMTAMPD	906
OY	664	-TIKYVEHPDERPHASNDGMGNASEHVLAKKDHSEDPKNKFADEPEVE-----ETPAE	716
Db	907	KTASGEETREVDHSPNNSKIGIQGTSEPQDEKQGS--PTEVIAVKRQPLETEVILINEQAPR	964
OY	717	PEVPOVELEKYEAOQKEAEVLLAKTYTDSGLKANATETTLAIGLRNNITQIOMDINSIMAEAE	776
Db	965	PEITEPGISKEETKMIENORFKETMETVLVAKAREGQ--LEVISKLTSHVKSEKLSHKK	1022
OY	777	KLIALILKGNPSVS	791
Db	1023	KTQIRRRASKPMVS	1037

```

RESULT 44
ID 081413 PRELIMINARY; .PRT; 1708 AA.
AC 081413;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Chromosome condensation protein, putative.
GN PFE0450W.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
NCBI_TaxID=36529;
RN RN SEQUENCE FROM N.A.
RP Devlin K., Baker S., Davies P., Mungal K., Berriman M., Pain A.,
RA Hall N., Bowman S., Churcher C., Quail M., Barrrell B.,
RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungal K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Chevarech I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corcoran C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden F., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gillwillam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphrey S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes K., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Mole S., Murphy L.,
RA Oliver K., Omond D., Price C., Quail M.A., Rabinowitz E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrrell B.G;
RL "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.",
RL Nature 419:527-531(2002).
OR EMBL; AL929352; CAD51456.1; -.
OR SEQUENCE 1708 AA; 20166 MW; 46A3F90A1437CE3B CRC64;

```

Dy

Query Match	4.0%;	Score 168.5; DB 5; Length 1708;
Best Local Similarity	17.6%; Pred.No. 0.62;	
Matches 177; Conservative 142; Mismatches 332; Indels 353; Gaps 39;		
24 KQATOKENTLTPDEVSKREGINABOIIVIKITDQGVTISHGD--HYHHYNGKVPYDAIIISE 81	:	: : : : :

Db	657	KEKOVHTIKYLTTEEBEKKNYLAKNO-----NDDEIOSKIEKLTNTLIEETIKENIKKE	711
QY	82	ELLMKDPN-----YKLKDEIVNEVG-----GYIVKVDGKYVYLKDAAH-----	122
Db	712	VVTDRNNMKLHQFIYNLEKK-----SKINGIGMLIDLYIEKKYEKAFITIASNNCSDEVVV	767
QY	123	---ADNVRTEKINRQOEHSQHHEGGTPRNDGAV-----ALASQGYTT	165
Db	768	ENPDVAVLFEVBK-----ANIGGVNVLSTVLNKLMTIMLKNEIYTQ	813
QY	166	DDGYIFNADSIIEPTGDAYIVPHGDHNYIPKNELASLELAEEAF-----LSG	214
Db	814	LLPRVYRLIDIKRNDKYK-----CYYIIKETLLANSLEQAHVIAVSHKKRVYTINGE	869
QY	215	-----KGNLSNSTRYPRONSDNSTRYWPVSVSPGTNTNTSNNSN	256
Db	870	LIENDGRI CGGMNDKWKVNGSNGS-----ERGINSRT-----STNNKNNNNNNHNNSN	920
QY	257	TNSQASQSN-----DIDSLKQLYLPLSQRIVESDGLVFPRAQTSTRTACVAVPHG	309
Db	921	SSSIKTSEYDESHLNSSEKTIKELNK-NIQDKKQKQDILINDIKDIN-----	967
QY	310	DHYHFIYPSQSEHEERIARIIPRYRSNH-----VPDSRPEPSPQPTPEPSPQPOP	363
Db	968	-----TFEDNECKIVIAKKRIDMKKQLEBIDDOLONSKTRPELTKEBENLNT	1016
QY	364	APNLKIDSNS-----LVSOLYRKVEGY-----	387
Db	1017	LKNLIEEKNEKSKVEIYVLKAOEKVKVYVEQLODVGEKKKLNKFNINAERQLTIMKD	1076
QY	388	-----VEEGISRYVPAK-401	
Db	1077	QLOEHTNEBANALASLEKSEKDIMSESNILEYANKELEIKITENGCAYEEVET	1136
QY	402	--DEPSETVNLJESKLSKQESVSHTTLTAKKENAVAPROEFYDXYANMLLTAEHAKLFPYNGK	459
Db	1137	LTNLNLIDQSIIEBKQKKQOVD-----ENISKOLEVNDLVYK-IEHLQKELNGYN	1188
QY	460	RNSPFO-----ALDKL-----ERLN---DESTNEKLVDD	487
Db	1189	KNEYONKIDEYMDLIKQSDKVIHNNMLSEWRCLRLLGKKKKINKIQOEDSTREDIMKD	1248
QY	488	LLAFAPITPERIGK-----PMSQIEYTE--DEVNIAQADKYTTSDDGIFD	533
Db	1249	-----EHDGKROQLDGDGEDGEVDEDEDEDEDDJGEEEB-----E	1290
QY	534	EHDIIISDEGAYVYPRHNGSHWIGKDSLSKEXKAAQAATYKEKQILPPSPDADVANKPTG	593
Db	1291	EEEBVEBEVEDMOEDMOEDMOYODVODVODVODGEBA-WREKNIIYHNEUYEDIKGNDPN	1349
QY	594	DSAAAIYNRVAKGEKRIFLVRPLYVTEHTEVEKONGNLIIPKHD---HYHNITKPAW---FD	646
Db	1350	DQ-----YDKLENQ-----IYLPNGRDHNVVTRNNHIGQDKSTLYHIMDKNKKRKLNEN	1400
QY	647	DHYTKANG-----YTLBFLPATIKYVYHPRERPRPSNGOMNASEHVLGKK	693
Db	1401	DTTTCPEBKKKVKKKKKLPDOLKDLLEBFGSDBELKEHNEYDIN--INDTDIDLNNKK	1458
QY	694	D-----HSEDPN-----KNFKADEPVEETPAEBEVPQVETEKVEAQ	730
Db	1459	DIETKLENKSLIEKKTPNLKIPQDYVNLKYDYKVVRRKQDVKKSGKEKDOKIKAAVYDSLCKN	1518
QY	731	LKEAEVLIAKYTDSLSKAAATETLAGLRNNLTLOIMDNNSIMAE	774
Db	1519	RKEEFVAFNVYSSKTK-EMYQWIAIGDAELEIIDSSIFPNE	1560

RESULT	45
O8IEB6	
ID O8IEB6	PRELIMINARY;
AC O8IEB6:	PRT; 2651 AA.
DT 01-MAR-2003	(TTEMBUrel. 23, Created)
DT 01-MAR-2003	(TTEMBUrel. 23, last sequence update)

DT 01-MAR-2003 (Tremblrel. 23, last annotation update)
 DE Ebl-1 protein.
 GN EBL-1.
 OS Plasmidium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmidium.
 NCBI_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Harris B., Lennard N., Clark L., Line A., Barron A., Corton C.,
 RA Berriman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,
 RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrett B.,
 RA Submitted (Sep-2002) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AL844509; CAD52344.1;
 SQ SEQUENCE 2651 AA; 305661 MW; 2702399D3979AA68 CRC64;

Query Match 4.0%; Score 166.5; DB 5; Length 2651;
 Best Local Similarity 18.7%; Pred. No. 1.5;
 Matches 180; Conservative 136; Mismatches 330; Indels 319; Gaps 44;

QY 14 ENNRVSYIDGKATQK-----TE-----NLTPDEVSKREGINAEQ 48
 DB 1171 ENNKGSHSDIKQDNDEGTDVSLTESEPKDLESVSFSSIDMDLKPKNKSFVTFDHEK 1230
 QY 49 IVIKITDGYVTSHGDIHYNGKVPYDAIISEELMKDPNYKLKDEDIYVE--VKGGYV 106
 DB 1231 ISVSPNVSVTVYDEGDK---RQGISDSSIHHEI---DPERNLHYESFISGLEGEGDI 1283
 QY 107 IKVDGKYVYL-----KDAADNVATKEIKRQKQKQSHREGGTPRDGVALAR 158
 DB 1284 EKEKKEGEGSLPISEPKINDGKRENVLDLSVSRVNDNRDISNVVSEEGESSIISSR 1343
 QY 159 --SOGRYTTDDGYINASDIIEDTDAYIVPHGDHYHYIPKUELASELAEAFLSGRG 216
 DB 1344 NGTEG-----INSEELKSEHTSYDVNNRDE--HNKQENLVSSSTQOSE----- 1387
 QY 217 NLSNSTRYRQNSDNTSRITWVPSVSNPDTTNTSNTSNTSQAQSNDI-----DSL 271
 DB 1388 ---RKRKRKEKAD--SHSELSLSISEVGET---IRRDAAEASENDKEDDIQCEEQQIV 1440
 QY 272 KQLYL-----PLSGRHYESDGLVFPDPAQITSRKRGVAVPHGDHYH----- 314
 DB 1441 TEKTLIEPTVNILOPSTPLGSHKES-----EIDDKONTGGEVHTDNTPTQYSS 1494
 QY 315 -----IPYSQMSLE-----ERIAIIPLRVRSNH-----W 340
 DB 1495 ESVEYITTKPDMENTTKPPSOYIKETIGEIDSTKQDNDECSNIIPIINKNNEEDGVS 1554
 QY 341 VPSDRPEQSPQPTPEPSPGPQAPNLIKIDNSNLSVS---QVRYKVGEGYFE----- 390
 DB 1555 IPRTR-----NVMESGFVSRNEQIIEKDKHITDDTTNPS 1592
 QY 391 EKGISRYVPAKDLBETVKNL---ESKLSKQDSVHTLTAK---KENAVPRQERYDA-- 443
 DB 1593 ENGLKGY---GEMPPDSIKSVITTESPLRDVQOMIEPIDGKNENKNIIIGEPQESTTEIR 1649
 QY 444 -----VNLTAHKALEFNKNRNSPQALDLRLRLNDESNKELVDLLAFLAP 494
 DB 1650 KOMDGPISNVNIPBELHPVAEGSK-----LESAKESMDDA--DKGTTIEDITV----- 1696
 QY 495 ITHPERLKPNQSLEYTEDEVRIALADKYTTSIDYIFDEHDIIDE-----GDYVYTP 548
 DB 1697 VEDPNCIGIEHONLKVHEQ---ASELNTYNSLDQRTNVEYERLDENPQGISPNRITTE 1752
 QY 549 HMGSHHWIGKDSLKEXVAAQAYYKEGILPPSPDADYKANFTDSDAAAIYNRKYEKR 608
 DB 1753 HI-----ELDKKEIHEPN---EIDAH--NGEOEMINNEVSDN-- 1786
 QY 609 IPIVRLPYMVEHTVEVKNGNLIIPKDHYHNKFKAMPFDHTYKANGTLLDLFLATIKY 668
 DB 1787 ---RMDSQISRDNTROLN---QDHESDSKDEIID-----RRE 1818
 QY 669 VEHPDERH--SNDGWNASSEHVLGKKDS-----EDPNKFKADEPEVEETPAP-- 717

DB 1819 MENLEENPNSSSSDLENPEGKEKGIETHSSSELDVDETYKXGIENQITETIEISV 1878
 QY 718 -----EVPQVTEKVEAQLKAEVLAQVDSILKA 748
 DB 1879 EQDNTIPGNSKETEVDSNRADMEBEKDVKIKQITIESE---BELEISKOTTSSHKEP 1935
 QY 749 NATE---TLAGLRNNLTLOIMDNN-----INAEKELALLKG-----SNPSSV 790
 DB 1936 NIEGQSVNIYDSKNEINQIKVQNEQNEGDPILLLEQKNIAILENQKNVNNPSQL 1995
 QY 791 S-KEX 794
 DB 1996 SHKER 2000

RESULT 46
 ID Q81223 PRELIMINARY; PRT; 1465 AA.
 AC Q81223;
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, last annotation update)
 DE Hypothetical protein (fragment).
 GN MALP3.1.
 OS Plasmidium falciparum (isolate 3D7).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmidium.
 NCBI_TaxID=36329;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99376085; PubMed=10448855;
 RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
 RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
 RA Gentles P., Gwilliam R., Hamlin N., Harris D., Holtroyd S., Hornsby T.,
 RA Horrocks P., Jagels K., Jasal B., Kyes S., McLean J., Moule S.,
 RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
 RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
 RA Whitehead S., Woodward J.R., Newbold C., Barrett B.G.;
 RT "The complete nucleotide sequence of chromosome 3 of Plasmidium falciparum".
 RL Nature 400:532-538 (1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22255708; PubMed=12368867;
 RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
 RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corton C.,
 RA Cronin A., Davies R., Davis P., Dear P., Dardien F., Doggett J.,
 RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
 RA Harper D., Hauser H., Hornsby T., Holtroyd S., Horrocks P.,
 RA Humphrey S., Jagels K., James K.D., Johnson D., Kerhornou A.,
 RA Knights A., Kontorov B., Kyes S., Larke N., Lawson D., Lennard N.,
 RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
 RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
 RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
 RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
 RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
 RA Sulston J.E., Craig A., Newbold C., Barrett B.G.;
 RT "Sequence of Plasmidium falciparum chromosomes 1, 3-9 and 13".
 RL Nature 419:527-531 (2002).
 DR EMBL; Z98547; CAD49084.1;
 KW Hypothetical protein.
 FT NON TER 1465 1465
 SQ SEQUENCE 1465 AA; 173879 MW; 9C50085A1BDE5B0 CRC64;

Query Match 4.0%; Score 166; DB 5; Length 1465;
 Best Local Similarity 18.6%; Pred. No. 0.7;
 Matches 170; Conservative 111; Mismatches 312; Indels 320; Gaps 42;

QY 12 VKNRNVSYIDGKATQKTEVLTPEDEVSKREGINAEQIVIT-----DQGYVTSHG 63
 DB 267 IKNNNEKALILNKNIVRVVINKKE-----EKNKKEFIRITQIRSMINEKNFORNED 322

```

QY 64 DHYHYNGKVPYDAIISEELMKDPNYKLDE-----DIVNEVKCG 104
DB 323 DHNNHKYDKKYEND---DHHNNHKYDKKYENDDDHHNNHKYDKKYDHDGHHNNHKYDKKYDHDG 379
QY 105 YVITKDGK-----YVILKDAADNVRTKEINROQES-----QREBG 146
DB 380 HHNNKYDKKYDSDGHHNNHKYDKKYDHDGHHNNKYD---KRYDGDGHHNNKYDKKYDHDG 436
QY 147 TPRND-----GAVALARQGRYTTDGG-----YIFNADIIEDTGDAYIVPH 188
DB 437 HHNNKYDKKYDGGHHNNHKYDKKYDHDGHHNNKYDHHNNKYDKKYDHDGHHNNKYDHDG 494
QY 189 GDHNYH-----IPKNELASASELAARFLSGRGN----- 217
DB 495 GEHVQHEETSDKNSPENNYSNNRSEGYILKRDVKLARVNSDKRVYKQJNSDEGYEN 554
QY 218 ---LSNSRTYRQNSDNTSRTWVPSVN--PGTTNTNTSNNNSNTNSQASQNSDIDSLK 272
DB 555 VLKYNTEKESYNSIDSIYREEDYREYNDVPSRYNDVNRYSNRDKKNEINENSYYK 614
QY 273 QL-----YKLPLSQRHVESDGLVFPDPAQITSTARGVAVPHGDHNFIF 317
DB 615 NVKYNKMFEMENDDKENEPEYAEVHYNDKYIDN--QYDDRYNG---NRDQYHFY-- 666
QY 318 SOMSELEERIAIIPLRYSNHWVPDSRPEOPSPOPTPEPSGQAPAPNL-----KIDSN 372
DB 667 ---DDEENVHK---SKYNNKYUNDYKDYNNKN-----AVNNYEEKOKLKY 708
QY 373 SSVLSQVLRKGE--GYVEEKGISRYVPAKDLPSFTVKNLESKLSKQESVSHLTAKKE 430
DB 709 SSFQKVDIKRGDENDYIYNNNN-----NNMDEKI-----HTLDKDH 746
QY 431 NVAPRDOEFYDKAVNLTAEHAKLFXNKGNSDFQALDKLERLNDESTNKEKLVDD--- 487
DB 747 NLDKNN--IYDDKNNIYDD-----KNNMYDDKNNIYVDKNNIYVDKNNMYDDKNN 794
QY 488 -----LAF LAP I-----THPERLQKPNQO---IETDEVARIAOLADKYTT- 526
DB 795 YGNNHIVGCTNTTYYLKNLPKSKYKRRESEGEINNESNNLSGEEYRMMQ--KKKNTTF 853
QY 527 -----SDGYFDEHDIISDEGDAYVTPHGSHWIGDLSLD----- 563
DB 854 KHIDENNMYKKNELINEYNTIDDPNLNNINNRNIIYMHKRNLSNEQNNKFNINNEL 913
QY 564 -----KEKVAQAAYTEKGIILPSPDADVKANPTGSAALYNR-----VKGEKRI 610
DB 914 EADEKKNNKMMFYNKERNYLINNOE-----YNNKYIYRNEYKNEKKNY 958
QY 611 LVRLPY-NVEH---TVEYKN-----GNLIIPHKDHYNIKF 642
DB 959 LDSYVEVYEHKNNKXTSMNNNOYDNKDSYLYDVSSSFNNKKYTKLI--KDNLCNNKY 1016
QY 643 AM--FDHT-----YKAPNGYLTEDLFATIKYVV--EHPDERPHSNDGNASEHV 689
DB 1017 KHNNDNNHNNIDGNDVYKGP--YGLEBGRAYDKVYHNDNLNDHLLYNDVNNYNTKVNI 1074
QY 690 LQKXHSRPNK 702
DB 1075 -----HEDHNN 1081

```

```

RN [1]
SEQUENCE FROM N.A.
RP STRAIN=3D7:
RX MEDLINE=99376085; PubMed=10448855;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltham T.,
RA Gentles S., Gilliam R., Hamlin N., Harris D., Holroyd S., Hornsby T.,
RA Horrocks P., Jagsels K., Jaseal B., Kyes S., McLean J., Moule S.,
RA Munnell K., Murphy L., Oliver K., Quail M.A., Rajadream M.A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrett B.G.,
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium
falciparum .".
RL Nature 400:532-538 (1999).
DR EMBL; AL034558; CAB39021.2;
KW Hypothetical protein.
SQ SEQUENCE 1826 AA; 216259 MW; 9E603E15FF29B276 CRC64;

Query Match 4.0%; Score 166; DB 5; Length 1826;
Best Local Similarity 18.6%; Pred. No. 0.95;
Matches 170; Conservative 111; Mismatches 312; Indels 320; Gaps 42;

QY 12 VKENNVASYIDGQATQKTEMLTPDEVSKREGINAEOIYIKIT-----DQGYVTSHG 63
DB 267 IKNNNEEKAIRLNKIYRVNINKEE---EKNNKESFIRITQIRSRMINEKNFORND 322
QY 64 DHYHYNGKVPYDAIISEELMKDPNYKLDE-----DIVNEVKCG 104
DB 323 DHNNHKYDKKYEND---DHHNNHKYDKKYENDDDHHNNHKYDKKYDHDGHHNNHKYDKKYDHDG 379
QY 105 YVITKDGK-----YVILKDAADNVRTKEINROQES-----QREBG 146
DB 380 HHNNKYDKKYDSDGHHNNHKYDKKYDHDGHHNNKYD---KRYDGDGHHNNKYDKKYDHDG 436
QY 147 TPRND-----GAVALARQGRYTTDGG-----YIFNADIIEDTGDAYIVPH 188
DB 437 HHNNKYDKKYDGGHHNNHKYDKKYDHDGHHNNKYDHHNNKYDKKYDHDGHHNNKYDHDG 494
QY 189 GDHNYH-----IPKNELASASELAARFLSGRGN----- 217
DB 495 GEHVQHEETSDKNSPENNYSNNRSEGYILKRDVKLARVNSDKRVYKQJNSDEGYEN 554
QY 218 ---LSNSRTYRQNSDNTSRTWVPSVN--PGTTNTNTSNNNSNTNSQASQNSDIDSLK 272
DB 555 VLKYNTEKESYNSIDSIYREEDYREYNDVPSRYNDVNRYSNRDKKNEINENSYYK 614
QY 273 QL-----YKLPLSQRHVESDGLVFPDPAQITSTARGVAVPHGDHNFIF 317
DB 615 NVKYNKMFEMENDDKENEPEYAEVHYNDKYIDN--QYDDRYNG---NRDQYHFY-- 666
QY 318 SOMSELEERIAIIPLRYSNHWVPDSRPEOPSPOPTPEPSGQAPAPNL-----KIDSN 372
DB 667 ---DDEENVHK---SKYNNKYUNDYKDYNNKN-----AVNNYEEKOKLKY 708
QY 373 SSVLSQVLRKGE--GYVEEKGISRYVPAKDLPSFTVKNLESKLSKQESVSHLTAKKE 430
DB 709 SSFQKVDIKRGDENDYIYNNNN-----NNMDEKI-----HTLDKDH 746
QY 431 NVAPRDOEFYDKAVNLTAEHAKLFXNKGNSDFQALDKLERLNDESTNKEKLVDD--- 487
DB 747 NLDKNN--IYDDKNNIYDD-----KNNMYDDKNNIYVDKNNIYVDKNNMYDDKNN 794
QY 488 -----LAF LAP I-----THPERLQKPNQO---IETDEVARIAOLADKYTT- 526
DB 795 YGNNHIVGCTNTTYYLKNLPKSKYKRRESEGEINNESNNLSGEEYRMMQ--KKKNTTF 853
QY 527 -----SDGYFDEHDIISDEGDAYVTPHGSHWIGDLSLD----- 563
DB 854 KHIDENNMYKKNELINEYNTIDDPNLNNINNRNIIYMHKRNLSNEQNNKFNINNEL 913
QY 564 -----KEKVAQAAYTEKGIILPSPDADVKANPTGSAALYNR-----VKGEKRI 610
DB 914 EADEKKNNKMMFYNKERNYLINNOE-----YNNKYIYRNEYKNEKKNY 958

```

```

QY 611 LVRLPY-NVEH-----TVEVKN-----GNLIIPKDHVHNKIF 642
DB 959 LDSYFVEWEHKKKTSMMNNNOYDKDSSYLYDYVSSSPNNKKYKTLII--KDNLCNNKY 1016
QY 643 AW--FDDHT-----YKAPNGYTLIEDLFAFIKYV--EHPDERPHSDGNGVASEHV 689
DB 1017 KHNNDNNNNIDGNDVYKHP--YGIEGRAYDKVYHNDNLNDHHLVNDVYNTKYNI 1074
QY 690 LGKXHDSEDPNKN 702
DB 1075 -----HEDHNNN 1081

RESULT 48
Q9NFS3 PRELIMINARY; PRT; 16215 AA.
AC Q9NFS3;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE D-titin.
GN SLS OR D-TITIN OR CG1915 OR CG18242 OR CG18245 OR CG18857.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OK NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RT Zhang Y.O., Broadie K.S.;
RL "Characterization of Drosophila D-titin gene."
CC Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
DR -i- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AJ271740; CAB93524.1; -.
DR HSSP; P56276; ITLK.
DR PDBase; F8gn0003432; s1s.
DR InterPro; IPR003961; FN_IIL.
DR InterPro; IPR007110; IG_Like.
DR InterPro; IPR003598; IG_C2.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00041; fn3; 5.
DR Pfam; PF00047; ig; 5.
DR Pfam; PF00018; SH3; 1.
DR SMART; SM00408; IGc2; 15.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50835; IG_Like; 50.
DR PROSITE; PS50002; SH3; 1.
KW Immunoglobulin domain; SH3 domain.
SQ SEQUENCE 16215 AA; 1841509 MM; 242C8765E00F7603 CRC64;

Query Match 4.0%; Score 166; DB 5; Length 16215;
Best Local Similarity 19.1%; Pred. No. 21;
Matches 189; Conservative 126; Mismatches 362; Indels 314; Gaps 43;

QY 24 KOATKENTLTPDEVSKE--GINABQIVIKI--TDQVYTHSGHYHYNKVYD 76
DB 4902 KSWTEITNLKATPIEKKKPAKVLKAKVLSIKITEROGIMSLGATLEQIAGKTEK 4961
QY 77 AIISEELMKDKNYKXK-----DEDIVNEKGVYKVDQKYVYVYKDAAH 122
DB 4962 AI--PWITRE--KLKAVESVOQLNKFDLDEVYLQPIEG--QIETESQ-----LPQHQ 5010
QY 123 ADVNTEKEINROKQESHQREGCTPRNDGAVALAROGGRYTTDDGYIFNSDIIE---- 178
DB 5011 VEOVQRTKEIQRLKSMESVEIEMETDQIDKLITQQ-----NAKDLIPWKEM 5057
QY 179 -----DTGDA-----YI 185
DB 5058 RQOLKSVORVTKQIDFKIEVELRHLQAOQAITEEYQGTAEFTVWIMDESSKGISKY 5117
QY 486 VPHGHYHYIPKNEISASELAAE-----AFLSGRNLNSRTRYRRQNDN-----TSRTNW 237

```

```

DB 5118 LRRDEQQLVEDQSNIVYKQFITTEDVNIIMHVSEKLEAQRLLREQAVVMRQQOQRPQL 5177
QY 238 VPSVSNGTITNTSNNNSNNSQAS-----OSNDIDSLKOLYK----- 276
DB 5178 QPLTSVEDTVLSQTSEROKLVQOQSFLTEAORQOFVQVEOSQMSLEEYEQKIINORTQ 5237
QY 277 -----LPLSORH-----VESDGLVFDAQITSRFAGVAV 306
DB 5238 QEAFSRQPREPQKFIVQEDSTLHLQERHDTQEQQLIQOQPVVMD-----RGRKK 5288
QY 307 PHGDHYHFIPIYSQMSLEBRIARIIPLYRSNH--WVPDSRPEQSP-----QPT 354
DB 5289 P--DQPVYQPOQBOVKAEEFVEK--PKTYEEMHDELVEPPIBQPOVPVMMERGKKKPPQ 5344
QY 355 P-----EPSPGQPPAP-----NLIKTSNSLSVQLV---RKVEGYVE 390
DB 5345 PQEXTFEEAHDELVEPPIVQOPEFPVPMWEGKKKVAQOETVLSQEVVQTSQVVEQOIVE 5404
QY 391 E--KGISRYVFAKDLPEETVYNLESKLSKQESVSHLTAKKENVAP-----RD 436
DB 5405 ETKTAARVRYVPRPEPQKVEQVTLKPTPRPRKAVKAEIQKPLRSTRPVQPVAE 5464
QY 437 QEFYDKAYNLTLE-----AHKALFXNKGNSDPQALDKLRLBNDESTNKEKLVDDLAF 492
DB 5465 QKAVEBATDELTEEPPIQPOQVMMERKKKP---QKPOEVTBIPPTLEIAVDTLSEEV 5520
QY 493 APIITHPE-----RLGKPNQIETEDVIRIQLADKYTTSQGYIFDEHDIISDE 541
DB 5521 PKPTEPQOPVPMARQOKKQKPEQKQELPKSLIIV-----DITEED 5564
QY 542 GDVAVTHMGSHMIGK-----DLSLDEKVA--AAAYTKEGILPPSPADYKAPPTG 593
DB 5565 LKRVQPEPQVLMERKKKRPQPODVLEKLDVAPTTYEKAVUDLPDEPVEKRPBP-- 5622
QY 594 DSAAIYNRVKGKRIPLVR-----LPYVNEHTVE--VKNGNLIIPKDHVHNKIFAMPD 646
DB 5623 ----VLMQR--GKKIKIPESTEVEHVDVAQIETVVKEDEMIVEKKRIKTK----- 5671
QY 647 DHTYKAPNGYTLIEDLFAFIKYVHEPDERPHSDGNGVASEHVYKGDHSDPKNKAKAD 706
DB 5672 --RPKSTKEVT--EELFE-----EQPEEISPEE--EVPQVEEIEEIEVEEKRRUKT 5720
QY 707 EEP-----VEETPAE-----PEVPQVEEYEAOLKEAEVLA--VTDSSLKANATE 752
DB 5721 KKPKLTQCVTEEBTEPHEIIEKESEVQOBEIYEKKKAKKPKTVAEKQLKEEIP 5780
QY 753 TLAGIRNNLTLOIMDNNSINAAEKLALALK 783
DB 5781 TEETVEEBETAE--DQQLVAESKKVKKVKK 5809

RESULT 49
Q917U4 PRELIMINARY; PRT; 18074 AA.
ID Q917U4
AC Q917U4;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE CG1915 protein.
GN SLS OR CG1915 OR CG18242 OR CG18245 OR CG18857.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OK NCBI_Taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RT STRAIN=Berkeley;
RC MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

```

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazewicz R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
 RA Abril J.F., Abmayyan A., An H.-J., Andrews-Pfankuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahle S., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes W., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Gary N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mlshina N.V., Mobery C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Paclet J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svitek R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng X.H.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celisner S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Barton J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Drenek D., Farfan D.,
 RA Ferrieres S., Frise E., Galle R.F., Gary N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Paclet J., Paragans V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svitek R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Mathews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celisner S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Seale S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celisner S.E., Gibbs R.A., Rubin G.M., Venter J.C.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Flybase; (SEP-2002) to the EMBL/GenBank/DBJ databases.
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 CC -i- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 DR EMBL, AE003473; AAC22226.2; -

DR HSPB, P56276; ITLK.
 DR Flybase; FBgn0003432; s1s.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR003599; I9.
 DR InterPro; IPR007110; I9-like.
 DR InterPro; IPR003598; I9_c2.
 DR InterPro; IPR003006; I9_MHC.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00041; fn3; 5.
 DR Pfam; PF00047; i5; 39.
 DR Pfam; PF00018; SH3; 1.
 DR SMART; SM00409; I9; 53.
 DR SMART; SM00408; I9c2; 46.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS00835; I9_LIKE; 50.
 DR PROSITE; PS00002; SH3; 1.
 DR Immunoglobulin domain; SH3 domain.
 KW SEQUENCE: 18074 AA; 2058295 MW; 0710C589B3B9D050 CRC64;

Query Match 4.0%; Score 166; DB 5; Length 18074;
 Best Local Similarity 19.1%; Pred. No. 24; Mismatches 362; Indels 314; Gaps 43;

Matches 189; Conservative 126; Mismatches 362; Indels 314; Gaps 43;

QY 24 KQATKNTNLPDEYSKRE--GINAEQIVIKI-----TDQGYVSHGHHYNGKVPYD 76
 DB 4902 KSMTEITNLKATPIEKKKPAKLEAAVVLKSTIEDDQGISLGLTLEQIINAKTERE 4961
 QY 77 AIISELLMKDENVLYLK-----DEDIVNEVKGQYVIVKDGKYYVYLKDAAH 122
 DB 4962 AI--FWITWRE--KLKAVESVQOQLNFKDLDDEVLYQLPLEG--QLETEGQ--LPQQAQ 5010
 QY 123 ANNVTKTEINQKQKHSHQRRGCGIPRNDGANALARSQGRYTTDDGCIYFNASDIIE---- 178
 DB 5011 VEQVQRTKIQRLKMSVESYIEMWTDQIDKITQGO-----NAKDLIPWKEM 5057
 QY 179 -----DTGDA-----YI 185
 DB 5058 RQQLKSVQRTVQIDKFKIEVEVLAHLQAQAITEEYQGTAEFVVMIDESSKGSISIV 5117
 QY 186 VPHGHHYHYPNLSLASELAAB--AFISGRNLNSRTYRQNSDN--TSRTNW 237
 DB 5118 LRRDQLQVEDQSNLYKKQKITTEEDVNIHVSEREKLEAQRILRQQAQVNRQOORPOL 5177
 QY 238 VSVSNPCTNTNTNNTSNNTSQAAS-----QNSDIDSLKQLYK----- 276
 DB 5178 QPLTIVEDTIVLSQTEROKLYVQOQSFTIEAQRQGFVQVEDSQMSLEIEHQKIINQRTQ 5237
 QY 277 -----LPLSORH-----VESDGLVDPDPAQITSRGAVAN 306
 DB 5238 QAFWRQPRQFQKTIQVEDSTLHLQERHDTQEOQLLOQCPVMD-----RGRKK 5288
 QY 307 PHGDHYHPIYVSQMSLEERLARIIPLYRSNH--WVPDSRQSP-----OPT 354
 DB 5289 P--DQPOVYQPOEQVKEEVEK--PKTYEEHMDLVEPTPEQOPVPMWVRGKKRQ 5344
 QY 355 P-----EPSPGPQAP-----NLKIDSNSLSVQSLV--RKYGSGIYVE 390
 DB 5345 PQEKTFEAHDELVEPTPVQOPEPVPVWVRGKKVVAQOETLSQEVVQTSQVVEQOIVE 5404
 QY 391 E--KGISRVFAKDPSETVKNLSKLSKQSVSHTLAKKENVAP-----RD 436
 DB 5405 ETKTKAVRVRVPPRPQEVQVETLKPTPRPKKAVNAEELQLKPLNSTRPQPVQVNAE 5464
 QY 437 QEFYDKAVNLTLE--AHKALFNXKGRNSDFQALDKLLERLNDSSTNKEKLVLDLAFV 492
 DB 5465 OKAYEANDDELTEEDIPQOPVPMWVRGKKR--QKQOEVEVTEIPKTLIAVDTLVEEV 5520
 QY 493 APITHPE-----RLGKPNQIYTBDEVVIAQLADKYTTSDGIYFPEHITLISE 541
 DB 5521 PKPTSPQOPVIMARGQKKPQKPDQKQELPKSLIAV-----DTIEED 5564
 QY 542 GDAYVTPHNGSHWIGK-----DSLSDKEVA--AQAYTKEKGLPPSPADAVANPFG 593

```

Db 5565 LIKVPQPEPQVPLMERKKKKKQPOQVIEBKLDVAETKYEKAVDVLDPDEKVEKPEP-- 5622
Qy 5564 DSAALIVRVGKERIPLVR-----LPMVEHTVE--VKNGLLIPHDHYNIFAMFD 646
Db 5623 ---VLMOR--GKKKIPKSEPTVEEVHPEDVDAQISTVKEDEMIYEERKRIKTKT----- 5671
Qy 647 DHTYKAPNGYLTLEDLFATIKYVEHPDERPHSNDGWNASEHVLGKQDSEDPNKNFAD 706
Db 5672 --RPSSTKEVT-EEJFE-----EQPEEISPEE--EVPQKEVIEIEIYEERKRLKKT 5720
Qy 707 EEP-----VEETPAE-----PEVPOVETEKVQAQKEAEVLLAK-VTDSILKANATE 752
Db 5721 KKPFLTQOVTEETPEEIIKESEEVQBOEIVEEKVKKKKTVAEKOLKEEIP 5780
Qy 753 TLGGRNNLTQIMDNNSIMAEKLLALK 783
Db 5781 TEEVVEEETAE--DQQLVEESKKVKVKVK 5809

RESULT 50
O813A6 PRELIMINARY; PRT; 1373 AA.
AC O813A8;
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DT 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Adapter-related protein, putative.
GN PF10200C.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxId=36329;
RN [1]
RP SEQUENCE FROM N. A.
RX MEDLINE=22255708; PubMed=12368867;
RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,
RA Mungall K., Bowman S., Atkin R., Baker S., Barron A., Brooks K.,
RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,
RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corcoran C.,
RA Cronin A., Davies R., Davis P., Dear P., Dearden R., Doggett J.,
RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,
RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,
RA Humphray S., Jagels K., James K.D., Johnson D., Kerhornou A.,
RA Knights A., Konfortov B., Kyes S., Larke N., Lawson D., Lennard N.,
RA Line A., Maddison M., McLean J., Mooney P., Moule S., Murphy L.,
RA Oliver K., Ormond D., Price C., Quail M.A., Rabinowitsch E.,
RA Rajandream M.A., Rutter S., Rutherford K.M., Sanders M., Simmonds M.,
RA Seeger K., Sharp S., Smith R., Squares R., Squares S., Stevens K.,
RA Taylor K., Tivey A., Unwin L., Whitehead S., Woodward J.,
RA Sulston J.E., Craig A., Newbold C., Barrall B.G.;
RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13.";
RL Nature 419:527-531(2002).
DR EMBL; AL929355; CAD51726.1;
SQ SEQUENCE 1373 AA; 159691 MW; C49205917D0695C2 CRC64;

Query Match 4.0%; Score 165.5; DB 5; Length 1373;
Best Local Similarity 17.9%; Pred. No. 0.68;
Matches 152; Conservative 143; Mismatches 293; Indels 261; Gaps 41;

Qy 35 PDEVS-----KREGINAEQIVIKITDQGYVTSHGDHYHYNGKV-----PY 75
Db 685 PDRVKSCLITAFKLCRRNNIRDHVAKKIIDK-----YKNSKITDLOQRCEY 733
Qy 76 DAISEELLMD---PNYKLR--DEDIVNEVKGYV---IKVNGKYV--YIKDAAH 122
Db 734 DLINNPILMNNVFSLRNTKKKIVIDEFL--SLNPYIEEHLRSGSKYSISKDLRKNQEH 791
Qy 123 ADVNRTKEI-----NROKQESHQREGTP-----RNDGAVALARSGRY 163
Db 792 DESIKSTPNVNTFTPELPNNKLSDDVNNISNNIPYEKNYVQEDAVVPMKATDEQA 851
Qy 164 TT--DDGYIF--NNSDIETDGDAYIVPHG--DHYHYIPKNELSASELAAAEAFISGRGN 217
Db 852 STMKDKGIYKLVNMGPKKWKKEVYNNIDGGQNDTDIAIEKRDMEKK---KKKSKGKN 907

```

```

Qy 218 LNSRTYRQNSDNTSRTNNVPSVSNPGTNTNTSNNSTNSQASQSDNDISLKKQLYKL 277
Db 908 NKKAKKEKNQNDNNNN-----NNNNNNNNNNNNHIDGNQNNSVINKDKYKK 954
Qy 278 PLSGRHVESDGLVDPQAQTSRTARGVAVPHGDHYHFI PYSQMSLEERIAIIPLRYS 337
Db 955 KKKKKKQRE-----EAADFN---KGIR-----NDINQTEEDI-----K 985
Qy 338 NHHVPDSRPEOPSPQPTPEBSPGQPPAPNLKIDSSLSQVLKVGEGYFEKIGISRY 397
Db 986 LHNIVNTQKR-----GMSNFS-----NYYDEEDDEDEY 1015
Qy 398 VFAPKDLPEFTYKQNESLKSQESVSHLTAKKENAVAPRODFY--DKAYNLLTEAHKALF 455
Db 1016 DDDDEDEBEYEDDEYEDDEEENEYVDESDDDEHDDDDYIEEOYN----- 1065
Qy 456 XNKGNSDFQALDKLERLNDSESTNKEKL-----VDDLAFAPITPHERLQKPSQIEYT 511
Db 1066 ---RNNVF-----INENENEKRTDVNVVDYNE-----QNYERF--DNRKDIHYS 1106
Qy 512 EDEVRIQADKYTTSQGYTFDEHDITSDGDAYVTTHMGHSHWI--GKDSLSEKELYAA 569
Db 1107 KEFNRF-----PSDNY-----HNKINNKTNNVNMKIKGINTVANNKKELTEKRWAA 1154
Qy 570 QAYTKEKGIIPSPDADVKNAPTGDSSAAATYNRVKGEKRIPLVRLPYMVEHTVEVKGNL 629
Db 1155 ALF---NGLISNN-----SPLKDPKSSY-----VSTIMKKQMS 1186
Qy 630 IIPKHDYHNIKFAFMDH---TYKAPNGYLTEDLFATIKYVEHPDERPHSNDGWNAS 686
Db 1187 ILSKNYFTSKKYKQNNNDMGIRKKSLLHNEQVERBDVT--KKKVNKGNKNNNNNDGCTNKL 1245
Qy 687 EHV-LGKQDSEDPNKNFK-----ADEEPVEETPAEPEVQVEFEKVEAQLKEAEVLLAK 740
Db 1246 VDIVNDSHNLNETTKKKNKNGDQDFMIDINELPKID---PKSSDINIEEEKKLWDK 1301
Qy 741 VTD-----SLLKANATELAGLRNNLTQIIMD-----NNS-----IMAEAE 776
Db 1302 ITTKKKAVFINSTYLNILQTLKKEINNINNTVIEVSKMDSLSCFYNNTKVLIKIKIEDN 1361
Qy 777 KLLALLKGS 785
Db 1362 KLIFLVKSN 1370

```

Search completed: November 14, 2003, 10:36:53
 Job time : 120 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 10:34:20 ; Search time 28 Seconds
(without alignments)
1202.837 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 4165

Sequence: 1 SYELGLYQARYTKENVRSY.....KLALLKGSNPSSVSKKXIN 796

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database :

1: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/prodata/2/iaa/5A_COMB.pep:*
4: /cgn2_6/prodata/2/iaa/5B_COMB.pep:*
5: /cgn2_6/prodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/prodata/2/iaa/backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4163	100.0	796	3	US-08-961-083-56
2	4163	100.0	796	4	US-09-536-784-56
3	4163	100.0	819	4	US-09-468-656A-8
4	2788.5	67.0	819	4	US-09-468-656A-10
5	2772	66.6	838	4	US-09-468-656A-4
6	2649.5	63.6	763	3	US-08-961-083-66
7	2649.5	63.6	763	4	US-09-536-784-66
8	1203.5	28.9	484	4	US-09-468-656A-6
9	1187.5	28.5	447	3	US-08-961-083-182
10	1187.5	28.5	447	4	US-09-536-784-182
11	163	3.9	3696	4	US-09-134-001C-5080
12	159	3.8	10182	4	US-09-134-001C-3159
13	147	3.5	812	4	US-09-513-783A-6
14	146	3.5	783	6	5231168-2
15	145.5	3.5	1073	3	US-09-541-782-6
16	145.5	3.5	1073	4	US-09-723-820-6
17	144	3.5	1040	4	US-08-961-083-118
18	144	3.5	1040	4	US-09-536-784-118
19	141.5	3.4	2314	4	US-09-816-703A-2
20	141	3.4	2431	1	US-07-920-281C-2
21	141	3.4	2431	4	US-08-466-277-2
22	141	3.4	8991	4	US-08-714-741-32
23	138	3.3	1503	3	US-08-976-255-14
24	137.5	3.3	1610	4	US-09-513-783A-22
25	136.5	3.3	1183	2	US-08-447-031A-2
26	136	3.3	812	4	US-09-513-783A-4
27	136	3.3	1056	4	US-09-513-783A-32

28	134	3.2	2308	1	US-08-015-973-1	Sequence 1, Appl
29	134	3.2	2308	2	US-08-448-164-1	Sequence 1, Appl
30	134	3.2	2308	3	US-08-081-929-2	Sequence 2, Appl
31	133.5	3.2	1664	1	US-09-599-652-2	Sequence 2, Appl
32	133.5	3.2	1664	2	US-08-642-846-2	Sequence 2, Appl
33	133.5	3.2	1664	4	US-09-264-604-2	Sequence 2, Appl
34	132.5	3.2	1507	6	5268270-2	Patent No. 5268270
35	132	3.2	1164	4	US-09-457-708-2	Sequence 2, Appl
36	132	3.2	1164	4	US-09-950-046A-2	Sequence 2, Appl
37	132	3.2	2032	4	US-09-071-033-462	Sequence 458, App
38	132	3.2	2032	4	US-09-071-033-462	Sequence 462, App
39	131	3.2	2032	4	US-09-071-033-466	Sequence 466, App
40	131	3.1	1093	5	PCT-US93-03077-1	Sequence 1, Appl
41	130.5	3.1	968	4	US-09-417-197-49	Sequence 49, Appl
42	130	3.1	2115	3	US-09-324-867-5	Sequence 5, Appl
43	129.5	3.1	525	4	US-09-107-532A-5095	Sequence 5095, Ap
44	129	3.1	2860	2	US-08-826-267-2	Sequence 2, Appl
45	128.5	3.1	916	1	US-08-188-228-48	Sequence 2, Appl
46	128.5	3.1	916	1	US-08-332-643-42	Sequence 48, Appl
47	128.5	3.1	916	1	US-08-332-638-48	Sequence 42, Appl
48	128.5	3.1	1848	5	US-08-296-791-6	Sequence 48, Appl
49	128.5	3.1	1848	5	PCT-US95-10661A-6	Sequence 6, Appl
50	128	3.1	783	4	US-09-513-783A-176	Sequence 6, Appl
51	127.5	3.1	2133	2	US-08-670-707A-37	Sequence 176, App
52	127.5	3.1	2133	3	US-09-037-601-37	Sequence 37, Appl
53	127.5	3.1	2133	4	US-09-315-179-37	Sequence 37, Appl
54	126.5	3.0	406	4	US-09-286-981B-18	Sequence 18, Appl
55	126.5	3.0	1181	2	US-08-488-940-2	Sequence 2, Appl
56	126	3.0	688	3	US-09-141-047-8	Sequence 8, Appl
57	125.5	3.0	2133	4	US-09-523-656-30	Sequence 30, Appl
58	125	3.0	868	1	US-07-864-004B-6	Sequence 6, Appl
59	125	3.0	868	1	US-08-251-937A-6	Sequence 6, Appl
60	125	3.0	868	1	US-08-212-133A-3	Sequence 3, Appl
61	125	3.0	1194	2	US-08-488-940-18	Sequence 18, Appl
62	125	3.0	2504	4	US-09-328-352-5821	Sequence 5821, Ap
63	124.5	3.0	1435	2	US-08-568-452A-4	Sequence 4, Appl
64	124.5	3.0	1435	2	US-08-487-826B-4	Sequence 4, Appl
65	124.5	3.0	1435	4	US-09-210-288-4	Sequence 4, Appl
66	124	3.0	635	4	US-09-417-197-135	Sequence 125, App
67	124	3.0	800	2	US-08-488-940-4	Sequence 4, Appl
68	124	3.0	813	2	US-08-488-940-3	Sequence 3, Appl
69	124	3.0	1194	2	US-08-488-940-1	Sequence 1, Appl
70	124	3.0	1194	4	US-08-488-940-17	Sequence 17, Appl
71	123.5	3.0	1601	4	US-09-345-473E-10	Sequence 40, Appl
72	123.5	3.0	2466	4	US-09-080-855-12	Sequence 12, Appl
73	123.5	3.0	2466	4	US-09-566-076-12	Sequence 12, Appl
74	123.5	3.0	2466	5	PCT-US94-09943-2	Sequence 2, Appl
75	123	3.0	628	3	US-08-656-034-2	Sequence 2, Appl
76	122.5	2.9	1338	1	US-08-471-033-50	Sequence 50, Appl
77	122.5	2.9	1338	2	US-08-471-044-50	Sequence 50, Appl
78	122.5	2.9	1338	2	US-08-463-048A-50	Sequence 50, Appl
79	122.5	2.9	1338	2	US-08-471-046A-50	Sequence 50, Appl
80	122.5	2.9	1338	2	US-08-470-566B-50	Sequence 50, Appl
81	122.5	2.9	1338	2	US-08-469-334-50	Sequence 50, Appl
82	122.5	2.9	1338	3	US-09-300-529-50	Sequence 50, Appl
83	122.5	2.9	1388	4	US-09-572-191-2	Sequence 2, Appl
84	122.5	2.9	1388	4	US-09-723-262-2	Sequence 2, Appl
85	122.5	2.9	1388	4	US-09-723-212-2	Sequence 2, Appl
86	122.5	2.9	2204	3	US-09-324-867-4	Sequence 4, Appl
87	122.5	2.9	2319	1	US-08-212-133A-8	Sequence 8, Appl
88	122.5	2.9	2319	1	US-08-474-503-6	Sequence 6, Appl
89	122.5	2.9	2319	2	US-08-670-707A-6	Sequence 6, Appl
90	122.5	2.9	2319	3	US-09-037-601-6	Sequence 6, Appl
91	122.5	2.9	2319	4	US-09-315-179-6	Sequence 6, Appl
92	122.5	2.9	2319	4	US-09-523-656-28	Sequence 28, Appl
93	122.5	2.9	2319	5	PCT-US94-13200-6	Sequence 6, Appl
94	122	2.9	1190	4	US-09-107-532A-7146	Sequence 7146, Ap
95	121.5	2.9	1177	4	US-09-134-001C-5106	Sequence 5106, Ap
96	121.5	2.9	2985	3	US-09-290-640-46	Sequence 46, Appl
97	121.5	2.9	2985	3	US-08-061-376-5	Sequence 5, Appl
98	121	2.9	542	1	US-07-814-964-13	Sequence 13, Appl
99	121	2.9	542	1	US-08-258-442-13	Sequence 13, Appl
100	121	2.9	542	1	US-08-328-809-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1

US-08-961-083-56

Sequence 56, Application US/08961083

Patent No. 6159469

GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/961,083

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8512

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 56:

SEQUENCE CHARACTERISTICS:

LENGTH: 796 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-961-083-56

Query Match 100.0%; Score 4163; DB 3; Length 796;

Best local similarity 100.0%; Pred. No. 1,38-314;

Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKNENRVSYIDGKQATQKTENLTPDEVSKREGINAEOIIVIKITDQGYVT 60

DB 1 SYELGLYQARTVKNENRVSYIDGKQATQKTENLTPDEVSKREGINAEOIIVIKITDQGYVT 60

QY 61 SHGHYHYHYNGKVPYDAIISEELMKDPYKDKEDINEVKGGVIVIKVDGKYVYLLDA 120

DB 61 SHGHYHYHYNGKVPYDAIISEELMKDPYKDKEDINEVKGGVIVIKVDGKYVYLLDA 120

QY 121 AHADVTKKEINROKQSHQREGGTPRNDGAVALARSGRYTTDDGYTFNASDIIDT 180

DB 121 AHADVTKKEINROKQSHQREGGTPRNDGAVALARSGRYTTDDGYTFNASDIIDT 180

QY 181 GDAYIVPHGDHYHYIPKNELSASELAABAFLSRGNSNSRTYRQNSDNTSRTNWVS 240

DB 181 GDAYIVPHGDHYHYIPKNELSASELAABAFLSRGNSNSRTYRQNSDNTSRTNWVS 240

QY 241 VSNPCTTNTNSNTNSQASQSDIOSLKLQYKPLSQRHVSDDLVPDPAQITSR 300

DB 241 VSNPCTTNTNSNTNSQASQSDIOSLKLQYKPLSQRHVSDDLVPDPAQITSR 300

QY 301 ARGVAVPHGDHYHYIPYSQMSLEERLARIIFLRYRSHNWVPDSRPEQSPQPTPEPSPG 360

DB 301 ARGVAVPHGDHYHYIPYSQMSLEERLARIIFLRYRSHNWVPDSRPEQSPQPTPEPSPG 360

QY 361 PQAPNLIKIDNSSLSVQLVKVGEGYVFEKKGISRYVFAADLPSEYKNELESTLSQES 420

DB 361 PQAPNLIKIDNSSLSVQLVKVGEGYVFEKKGISRYVFAADLPSEYKNELESTLSQES 420

QY 421 VSHLTAKKENVAPRDOEFYDKAYNLLTEAHKALFXKNGRNSDQALDKLERINDESTN 480

DB 421 VSHLTAKKENVAPRDOEFYDKAYNLLTEAHKALFXKNGRNSDQALDKLERINDESTN 480

QY 481 KEKLVDDLAFIATIPHERKGPNSQIETDEVRIAQLADKYTSDGYIFDEHDIIISD 540

DB 481 KEKLVDDLAFIATIPHERKGPNSQIETDEVRIAQLADKYTSDGYIFDEHDIIISD 540

QY 541 EGDAYVTPHMGSHWIGKDSISDKEXAAQAYTEKGIPLPSPPADVYKANTGSSAAIY 600

DB 541 EGDAYVTPHMGSHWIGKDSISDKEXAAQAYTEKGIPLPSPPADVYKANTGSSAAIY 600

QY 601 NRVGKERIPLVRPLPVWEHTVEYKNGNLIIPKHQYHNIKFAWFDHTYKAPNGYTLSD 660

DB 601 NRVGKERIPLVRPLPVWEHTVEYKNGNLIIPKHQYHNIKFAWFDHTYKAPNGYTLSD 660

QY 661 LFATIKYVEHPDERPHSNDGMSNAHEVLGKDHSEDPNKNFKADESPVEETPAEPEVP 720

DB 661 LFATIKYVEHPDERPHSNDGMSNAHEVLGKDHSEDPNKNFKADESPVEETPAEPEVP 720

QY 721 QVETEKYBAQLKEAEVLLAKYTSSLKANATETLAGRNLTLOIMONNSIMAAEAKLLA 780

DB 721 QVETEKYBAQLKEAEVLLAKYTSSLKANATETLAGRNLTLOIMONNSIMAAEAKLLA 780

QY 781 LKGSNPSVSKEKIN 796

DB 781 LKGSNPSVSKEKIN 796

RESULT 2

US-09-536-784-56

Sequence 56, Application US/09536784

Patent No. 6573082

GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/536,784

FILING DATE: 30-Oct-1997

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/961,083

FILING DATE: OCT-30-1997

ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks

REGISTRATION NUMBER: 41,971

REFERENCE/DOCKET NUMBER: PB340P3

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8512

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 56:

SEQUENCE CHARACTERISTICS:

LENGTH: 796 amino acids

TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 56:
 US-09-536-784-56

Query Match 100.0%; Score 4163; DB 4; Length 796;
 Best Local Similarity 100.0%; Pred. No. 1,3e-314;
 Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYELGLYQARTYKNNRVSYIDGKATQKTENLTPEVSKREGINAEQIVIKITTDGYYT 60
 DB 1 SYELGLYQARTYKNNRVSYIDGKATQKTENLTPEVSKREGINAEQIVIKITTDGYYT 60
 QY 61 SHGDHYYNGKVPDAIISEELMKDPYKLDKEDIINEVGKGYIKVDGKYVYVLLKDA 120
 DB 61 SHGDHYYNGKVPDAIISEELMKDPYKLDKEDIINEVGKGYIKVDGKYVYVLLKDA 120
 QY 121 AAADVNRKKEINRQKQHSQHRREGTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 180
 DB 121 AAADVNRKKEINRQKQHSQHRREGTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 180
 QY 181 GDAYIVPHGDHYIYIPKNEISASELAEEAFISGRGNLSNSTRYRONSNTSRITWVPS 240
 DB 181 GDAYIVPHGDHYIYIPKNEISASELAEEAFISGRGNLSNSTRYRONSNTSRITWVPS 240
 QY 241 VSNPGTTNTNTSNSTNSQASQNDISLKLQYKLPISQHRVESDGLVFPDAQITSTR 300
 DB 241 VSNPGTTNTNTSNSTNSQASQNDISLKLQYKLPISQHRVESDGLVFPDAQITSTR 300
 QY 301 ARGVAVPHGDHYHFIYPSQMSLEERIIARIIPLYRSNMHWPDSPREQSPQPTPEPSPG 360
 DB 301 ARGVAVPHGDHYHFIYPSQMSLEERIIARIIPLYRSNMHWPDSPREQSPQPTPEPSPG 360
 QY 361 POPAPVLTIDSNSLSVQLVRKVGEGYVEEEKISRYVFAKDLPSRTVKNLESKLSKQES 420
 DB 361 POPAPVLTIDSNSLSVQLVRKVGEGYVEEEKISRYVFAKDLPSRTVKNLESKLSKQES 420
 QY 421 VSHITLAKKENVAPRDOEFYDKAYNLLTEAHKALFNKGRNSDFQALDKLERLNDESTN 480
 DB 421 VSHITLAKKENVAPRDOEFYDKAYNLLTEAHKALFNKGRNSDFQALDKLERLNDESTN 480
 QY 481 KEKLVDDLLAFAPITHPERLGRKPSQIETDEDEVRIAOLADKYTTSDGYIFDEHDIISD 540
 DB 481 KEKLVDDLLAFAPITHPERLGRKPSQIETDEDEVRIAOLADKYTTSDGYIFDEHDIISD 540
 QY 541 EGDAYVTPMHGSHWIGKDSLSDKEKVAQAAYTKEKGIIPSPDADVKANPTGDSAAAY 600
 DB 541 EGDAYVTPMHGSHWIGKDSLSDKEKVAQAAYTKEKGIIPSPDADVKANPTGDSAAAY 600
 QY 601 NRKVGKRIPLVRLPVMVEHTVEVKNGNLIIPKHQHYHNIKRAWPDHYYKAPNGTLLD 660
 DB 601 NRKVGKRIPLVRLPVMVEHTVEVKNGNLIIPKHQHYHNIKRAWPDHYYKAPNGTLLD 660
 QY 661 LFATIKYYVEHDERPHSNDGWNASEHVLGKKDSEDPKNKPKADEEVEETPAPEVP 720
 DB 661 LFATIKYYVEHDERPHSNDGWNASEHVLGKKDSEDPKNKPKADEEVEETPAPEVP 720
 QY 721 QVETEKVEAQLKEAEVLLAKVTDSISKANATETLAQRNNLTLOIMDNNISMAEAEKLLA 780
 DB 721 QVETEKVEAQLKEAEVLLAKVTDSISKANATETLAQRNNLTLOIMDNNISMAEAEKLLA 780
 QY 781 LKGSNPSVSKEKIN 796
 DB 781 LKGSNPSVSKEKIN 796

RESULT 3
 US-09-468-656A-8
 ; Sequence 8, Application US/09468656A
 ; Patent No. 6582706
 ; GENERAL INFORMATION:

APPLICANT: Johnson, Leslie S.
 APPLICANT: Adamou, John E.
 TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
 TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
 TITLE OF INVENTION: Motifs
 FILE REFERENCE: 469201-444
 CURRENT APPLICATION NUMBER: US/09/468,656A
 PRIOR FILING DATE: 1999-12-02
 PRIOR APPLICATION NUMBER: 60/113,048
 NUMBER OF SEQ ID NOS: 14
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 8
 TYPE: PR1
 LENGTH: 819
 ORGANISM: Streptococcus pneumoniae
 US-09-468-656A-8

Query Match 100.0%; Score 4163; DB 4; Length 819;
 Best Local Similarity 99.9%; Pred. No. 1,4e-314;
 Matches 795; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SYELGLYQARTYKNNRVSYIDGKATQKTENLTPEVSKREGINAEQIVIKITTDGYYT 60
 DB 21 SYELGLYQARTYKNNRVSYIDGKATQKTENLTPEVSKREGINAEQIVIKITTDGYYT 80
 QY 61 SHGDHYYNGKVPDAIISEELMKDPYKLDKEDIINEVGKGYIKVDGKYVYVLLKDA 120
 DB 81 SHGDHYYNGKVPDAIISEELMKDPYKLDKEDIINEVGKGYIKVDGKYVYVLLKDA 140
 QY 121 AAADVNRKKEINRQKQHSQHRREGTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 180
 DB 141 AAADVNRKKEINRQKQHSQHRREGTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 200
 QY 181 GDAYIVPHGDHYIYIPKNEISASELAEEAFISGRGNLSNSTRYRONSNTSRITWVPS 240
 DB 201 GDAYIVPHGDHYIYIPKNEISASELAEEAFISGRGNLSNSTRYRONSNTSRITWVPS 260
 QY 241 VSNPGTTNTNTSNSTNSQASQNDISLKLQYKLPISQHRVESDGLVFPDAQITSTR 300
 DB 261 VSNPGTTNTNTSNSTNSQASQNDISLKLQYKLPISQHRVESDGLVFPDAQITSTR 320
 QY 301 ARGVAVPHGDHYHFIYPSQMSLEERIIARIIPLYRSNMHWPDSPREQSPQPTPEPSPG 360
 DB 321 ARGVAVPHGDHYHFIYPSQMSLEERIIARIIPLYRSNMHWPDSPREQSPQPTPEPSPG 380
 QY 361 POPAPVLTIDSNSLSVQLVRKVGEGYVEEEKISRYVFAKDLPSRTVKNLESKLSKQES 420
 DB 381 POPAPVLTIDSNSLSVQLVRKVGEGYVEEEKISRYVFAKDLPSRTVKNLESKLSKQES 440
 QY 421 VSHITLAKKENVAPRDOEFYDKAYNLLTEAHKALFNKGRNSDFQALDKLERLNDESTN 480
 DB 441 VSHITLAKKENVAPRDOEFYDKAYNLLTEAHKALFNKGRNSDFQALDKLERLNDESTN 500
 QY 481 KEKLVDDLLAFAPITHPERLGRKPSQIETDEDEVRIAOLADKYTTSDGYIFDEHDIISD 540
 DB 501 KEKLVDDLLAFAPITHPERLGRKPSQIETDEDEVRIAOLADKYTTSDGYIFDEHDIISD 560
 QY 541 EGDAYVTPMHGSHWIGKDSLSDKEKVAQAAYTKEKGIIPSPDADVKANPTGDSAAAY 600
 DB 561 EGDAYVTPMHGSHWIGKDSLSDKEKVAQAAYTKEKGIIPSPDADVKANPTGDSAAAY 620
 QY 601 NRKVGKRIPLVRLPVMVEHTVEVKNGNLIIPKHQHYHNIKRAWPDHYYKAPNGTLLD 660
 DB 621 NRKVGKRIPLVRLPVMVEHTVEVKNGNLIIPKHQHYHNIKRAWPDHYYKAPNGTLLD 680
 QY 661 LFATIKYYVEHDERPHSNDGWNASEHVLGKKDSEDPKNKPKADEEVEETPAPEVP 720
 DB 681 LFATIKYYVEHDERPHSNDGWNASEHVLGKKDSEDPKNKPKADEEVEETPAPEVP 740
 QY 721 QVETEKVEAQLKEAEVLLAKVTDSISKANATETLAQRNNLTLOIMDNNISMAEAEKLLA 780
 DB 741 QVETEKVEAQLKEAEVLLAKVTDSISKANATETLAQRNNLTLOIMDNNISMAEAEKLLA 800

```

QY 781 LKGSNPSSVSKEKIN 796
Db 801 LKGSNPSSVSKEKIN 816

RESULT 4
US-09-468-656A-10
; Sequence 10, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John B.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; CURRENT FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; LENGTH: 819
; SEQ ID NO 10
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-10

Query Match
Best Local Similarity 67.0%; Score 2788.5; DB 4; Length 819;
Matches 544; Conservative 85; Mismatches 139; Indels 47; Gaps 7;

QY 1 SYELGLYQA-RTVKENRVSYYDQKQATOKTENLTPEVSKREGINAEQIVIKITDQGV 59
Db 21 SYELGRYQAGQDKKESNRVAYIDQAGQKAENLTPEVSKREGINAEQIVIKITDQGV 80
QY 60 TSHGHHYHYNGKVPYDAIIESEELMKDPNYKLKDEDI VNEVKGGYVIXVKGKYYVYKLD 119
Db 81 TSHGHHYHYNGKVPYDAIIESEELMKDPNYQLKDSI VNEIKGGYVIXVKGKYYVYKLD 140
QY 120 AAHADNVRTKEEINRQKQESHQREGGTPRNDGAVALARSGRYTTDDGYIFNADIIED 179
Db 141 AAHADNIRTKKEIKRQKQESHNNHNS---RADNAVAARAQGRYTTDDGYIFNADIIED 197
QY 160 TGDAYIVPHGHHYIIPKNEISASBELAAEAFLSGRGLNSRRTYRRQNSDNTSRTNWP 239
Db 198 TGDAYIVPHGHHYIIPKNEISASELAAEAAYWNG-----KQSRPSSSSSYNA 246
QY 240 SVSNNGTNTNTSNNSNTNSQASQNDIDSLKQLYKLPLSQRHVESDGLVFPDQAQITSR 299
Db 247 NPAQRILSENHNLVTPTPTPHQ--NQGENISSILRELYAKPLSERHVESDGLVFPDQAQITSR 305
QY 300 TARGAIVPHGHHYIIPYQMSLEERLARIIPLYRSNHWPDSRPEQSPQPTPEPSP 359
Db 306 TARGAIVPHGHHYIIPYQMSLEERLARIIPLYRSNHWPDSRPEQSPQPTPEPSP 365
QY 360 GPQAPNLKIDSNSLSVQLVRKVBGVYFEEKGISRYVFAKDLPSFTVKNLESKLSQOE 419
Db 366 SPQAPNSPID--GKLVKAVRKVDGVEFENGVSRYTPAKDLAETAAIGDSKLAQOE 423
QY 420 SVSHLTAKKENAVARPDQFYDAKYNLLTEAHKALFXNKGNSDQALDKLERLNDEST 479
Db 424 SLSHLTGKTCTDLPSSDREFYKAYDLARIHQDLIDNKGQVDFEALDNLLERLKDVSS 483
QY 480 NKEKLVDDLAPLAPITPBERLQKPNQJETEDEVIRIQAQADKYYTSDGYIFDEHDITS 539
Db 484 DKYKLVEDLAPLAPITPBERLQKPNQJETEDEVIRIQAQADKYYTSDGYIFDEHDITS 543
QY 540 DEGDAYVTPHMGHSHWIGKQSLSDKEKYAAQAYTEKGLTPSPADAVKANFTGDSAAAI 599
Db 544 DEGDAYVTPHMGHSHWIGKQSLSEAEKRAAQAAYTEKGLTPSPSTHODSGNTEAKGAEXI 603

```

```

QY 600 YNRVKGKRIPLVRLPYVHEHTVEVKNQNLIIIPKHDYHNIKFAFPDHTYKANGYTL 659
Db 604 YNRVKAQKVPDMDPRLNQLQYTVVEVKNQSLIIPHYDHYHNIKFEFPEGLYEAQKGTLE 663
QY 660 DLPAITKYVHEHPDERPHSDNGKASBHVLGKKDQSHEDPNKFKADE-----PVEET 713
Db 664 DLATVKKYVHEHPDERPHSDNGKASBHVLRNKGQADDTQTEKPEEEKQTEKPEEBET 723
QY 714 -----PAPEVPOVETEKYEAOLKEAELLAKVTDSLSKANA 750
Db 724 PREEKPESEKPEKPTPEEPSESEBEPQVETEKVEKUREADLLGKIQDPTIKSNA 783
QY 751 TETLAGLNNLTLLQIMDNNSIMAEAKLLALKGS 785
Db 784 KETITGLKNNLTFGTQDNTTMAEAELKLLALKS 818

RESULT 5
US-09-468-656A-4
; Sequence 4, Application US/09468656A
; Patent No. 6582706
; GENERAL INFORMATION:
; APPLICANT: Johnson, Leslie S.
; APPLICANT: Adamou, John B.
; TITLE OF INVENTION: Vaccine Compositions Comprising Streptococcus
; TITLE OF INVENTION: Pneumoniae Group A and B Having Selected Structural
; TITLE OF INVENTION: Motifs
; FILE REFERENCE: 469201-444
; CURRENT APPLICATION NUMBER: US/09/468,656A
; CURRENT FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: 60/113,048
; PRIOR FILING DATE: 1998-12-21
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-468-656A-4

Query Match
Best Local Similarity 66.6%; Score 2772; DB 4; Length 838;
Matches 541; Conservative 91; Mismatches 144; Indels 56; Gaps 8;

QY 1 SYELGLYQARTV-KENNRVSYYDQKQATOKTENLTPEVSKREGINAEQIVIKITDQGV 59
Db 21 SYELGRYQAGQDKKESNRVAYIDQAGQKAENLTPEVSKREGINAEQIVIKITDQGV 80
QY 60 TSHGHHYHYNGKVPYDAIIESEELMKDPNYKLKDEDI VNEVKGGYVIXVKGKYYVYKLD 119
Db 81 TSHGHHYHYNGKVPYDAIIESEELMKDPNYQLKDSI VNEIKGGYVIXVKGKYYVYKLD 140
QY 120 AAHADNVRTKEEINRQKQESHQREGGTPRNDGAVALARSGRYTTDDGYIFNADIIED 179
Db 141 AAHADNIRTKKEIKRQKQESHNNHNGGS--NDQAVVAARAQGRYTTDDGYIFNADIIED 198
QY 180 TGDAYIVPHGHHYIIPKNEISASBELAAEAFLSGRGLNSRRTYRRQNSDNTSRTNWP 239
Db 199 TGDAYIVPHGHHYIIPKNEISASELAAEAAYWNG-----KQSRPSSSSSYNA 247
QY 240 SVSNNGTNTNTSNNSNTNSQASQNDIDSLKQLYKLPLSQRHVESDGLVFPDQAQITSR 299
Db 248 NPAQRILSENHNLVTPTPTPHQ--NQGENISSILRELYAKPLSERHVESDGLVFPDQAQITSR 306
QY 300 TARGAIVPHGHHYIIPYQMSLEERLARIIPLYRSNHWPDSRPEQSPQPTPEPSP 359
Db 307 TARGAIVPHGHHYIIPYQMSLEERLARIIPLYRSNHWPDSRPEQSPQPTPEPSP 366
QY 360 GPQAPNLK-IDSN---SLVQLVRKVBGVYFEEKGISRYVFAKDLPSFTVKNLESKLSQ 415
Db 367 SPQAPNSPID--GKLVKAVRKVDGVEFENGVSRYTPAKDLAETAAIGDSKLAQOE 426
QY 416 SKQSVSHTLTAKKENAVARPDQFYDAKYNLLTEAHKALFXNKGNSDQALDKLERLN 475

```

```

Db      427 AKQESLSHKLGAQKTDLPSSDREFYNKAYDILLARIHODLLDNKGRQVDFEALDNLLERLK 486
      476 DESTNKEKLVDDLLAFAPITPHERLQKPNQOIEYTEDEVRIAQIADKYTSDGYIFDEH 535
      487 DVPSDKVLKVDLAFAPIRPHERLQKPNQOITTYDDEIQAVALGKTYTDEGYIFDPR 546
Qy      536 DIISDEGDAVYTPHMGSHWIGKDSLDKEKVAQAQYTEKGIPLPSPADVANKPTGDS 595
      547 DITSDEGDAVYTPHMGSHWIKKDSLSEARAAQYAAKEKGLTPSTDHQDSGNTBAGK 606
Qy      596 AAATYRVKGEKRIPLVRLPYVEHTVEYKNGLLIPKHQYHNKIPAFMDHTYKAPNG 655
      607 ABAIYRVKAAKVPPLDRMPYNLQYTVVEYKNGSLIIPHYDHNKIKFEWDEGLYAPK 666
Qy      656 YTLIEDLFAITKYVVEHPDERPHSDMGNASEHVLRK-----KHSE----- 697
      667 YTLIEDLLATVKKYVVEHPNERPHSDNGFNASDHRKNKYDQSKPEDKEHDEVPETHP 726
Qy      698 -----DPRNFKADEEPVEETPAPEPPOVETEKYEAOLKEAVAL 738
      727 ESDEKENHAGLNPSADNLVYKPSDTDETEEEADTTDEAEIIPQVENSVINAKIADAEALL 786
Db      739 AKYDSSGLKANATETLAGLRNNTLTQIMDNNSIMAEAEKLLALLKGSNPSSV 790
      787 EKVTDPISIRQNAEMETLTGLKSSLLTGKNNNTISAEDVSLALLKESQAPAPI 838

```

RESULT 6

```

US-08-961-083-66
; Sequence 66, Application US/08961083
; Patent No. 6159469
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ. ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-961-083-66

```

```

Query Match      63.6%; Score 2649.5; DB 3; Length 763;
Best Local Similarity 66.8%; Pred. No. 4e-197;

```

```

Matches 519; Conservative 77; Mismatches 132; Indels 49; Gaps 8;
Qy      1 SYELGLQARVY-KENNRVSYIDGQATQKTNLPDEVSXKEGINAEQIVIKITQGVY 59
      2 STELGRHQGVQYKESNRVSYIDGQAKAENLPDEVSXKEGINAEQIVIKITQGVY 61
Qy      60 TSHGDHYHYNKKVPYDAIISEELMKDPNYLTKDEDIVNEVYKGVIVYDQYVYVYLD 119
      62 TSHGDHYHYNKKVPYDAIISEELMKDPNYLTKDEDIVNEVYKGVIVYDQYVYVYLD 121
Db      120 AAHADNVRTKEEINRQKQESHQREGGTPRNDCAVALARSQGRYTTDDGYIFNADIID 179
      122 AAHADNIRTKKEIRKQKQESHQREGGTPRNDCAVALARSQGRYTTDDGYIFNADIID 178
Qy      180 TGDAYIVPHGDHYHYIPKVELSASELAALASLGRGNSNRRTYRQNSDTSRNNWP 239
      179 TGDAYIVPHGDHYHYIPKVELSASELAALASLGRGNSNRRTYRQNSDTSRNNWP 227
Qy      240 SVSNPGTNTNTSNNSNTNSQASQSDIDSLKQLYKPLSORHVESDGLVFPDPAQITSR 299
      228 NPAQRLSENHNLVTPTTHQ-NQGENISSLLRELAKLSRHSVSDGLIFDPAQITSR 286
Qy      300 TARGVAVPHGDHYHYIPYSQSELEERARITPLRYRSNHWYDPSRPEQSPQTPPEPSP 359
      287 TARGVAVPHGDHYHYIPYSQSELEERARITPLRYRSNHWYDPSRPEQSPQTPPEPSP 346
Qy      360 GPQAPNPK-IDSN---SLVSQLVKVGEGVFEKGSRYVFAKDLPSRYKNLESL 415
      347 SPQAPNPPQAPSNPIDKLVKAVRKVGDGVFEEENGVSRYIPAKDLSEATAAGIDSKL 406
Qy      416 SKQESVSHLTAKKENVAPRDOEFYDKAYNLTEAHKALFXNKGNSDFOALDKLERLN 475
      407 AKQESLSHKLGAQKTDLPSSDREFYNKAYDILLARIHODLLDNKGRQVDFEALDNLLERLK 466
Qy      476 DESTNKEKLVDDLLAFAPITPHERLQKPNQOIEYTEDEVRIAQIADKYTSDGYIFDEH 535
      467 DVPSDKVLKVDLAFAPIRPHERLQKPNQOITTYDDEIQAVALGKTYTDEGYIFDPR 526
Db      536 DIISDEGDAVYTPHMGSHWIGKDSLDKEKVAQAQYTEKGIPLPSPADVANKPTGDS 595
      527 DITSDEGDAVYTPHMGSHWIKKDSLSEARAAQYAAKEKGLTPSTDHQDSGNTBAGK 586
Qy      596 AAATYRVKGEKRIPLVRLPYVEHTVEYKNGLLIPKHQYHNKIPAFMDHTYKAPNG 655
      587 ABAIYRVKAAKVPPLDRMPYNLQYTVVEYKNGSLIIPHYDHNKIKFEWDEGLYAPK 646
Qy      656 YTLIEDLFAITKYVVEHPDERPHSDMGNASEHVLRK-----P 709
      647 YTLIEDLLATVKKYVVEHPNERPHSDNGFNASDHRKNKYDQSKPEDKEHDEVPETHP 706
Qy      710 VEET-----PAPEVPOVETEKYEAOLKEAVALAKYTD 743
      707 EETPREKPKQESKESPRTPEBPSPESEPOVETEKYEAOLKEAVALAKYTD 763

```

RESULT 7

```

US-09-536-784-66
; Sequence 66, Application US/09536784
; Patent No. 6573082
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33

```


QY 541 EGDVYTPHMGSHWIGKDSLDKE 565
DB 453 DESGFVMSHGDHNYFFKKDLTEBQ 477

RESULT 9

US-08-961-083-182
Sequence 182, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 182:
SEQUENCE CHARACTERISTICS:
LENGTH: 447 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-182

Query Match 28.5%; Score 1187.5; DB 3; Length 447;
Best Local Similarity 46.3%; Pred. No. 5.5e-84;
Matches 257; Conservative 72; Mismatches 117; Indels 109; Gaps 10;
QY 4 LGLVARTVKNRRSYIDGKQATKNTLTPDEVSKREGINAEOIVIKITDQGYVTSHG 63
DB 1 LNQHRSOENKDNRRSYVDGSSQSKSENLTPDQVSQKREGIOAEQIVIKITDQGYVTSHG 60
QY 64 DHYHYNGKVPYDALISELLMKDPNYKDKEDIVNEVGVIVKVDGKYVYLLDAHA 123
DB 61 DHYHYNGKVPYDALISELLMKDPNYKDKEDIVNEVGVIVKVDGKYVYLLDAHA 120
QY 124 DNVRTKEEINROKQESHQREGGTFRNDGAVALLARSQGRYTTDDGIYFNASDIIEDTGA 183
DB 121 DNVRTKDEINROKQEHVKN-----KVNQNVAVASQGRYTTNDGIVFNPADIIEDTGA 176
QY 184 YIVPHGDHYHYPKNELSASELAAEAFLSGRGNLSNRTYRRONSNDNTRTNWVPSVN 243
DB 177 YIVPHGCHYHYPKSDLSASELAAKAHLAGKMQPSQLSYSTASDN----- 224
QY 244 PGTNTNNSNTNSQASQSDIDSLKQYKLPISQGHVSDGIVFPAPQITSTIANG 303
DB 225 ---NTQSAAKSTKPAKSNENLOSLLKELVDSPAQGVSSSDGIVFPAPKISTTPNG 280
QY 304 VAVPHGDHYHYPYQMSLEERIRIILPFRYSNHWVDSRPEQSPQPTPEPSPQPP 363

DB 281 VAIIPHGDHYHIFIPYSKLSALEEKIARWP----- 309
QY 364 APNLKIDNSSLVQVAKVGBGVYFEKGISRYFADLPSETVKNLESKSKQESVSH 423
DB 310 -----ISGTGSTVSTNAR-----PNEVVSLSGSLSSNPS- 339
QY 424 TLTKAKENVAPRDEQFYDKAVVLLTEAHKALFXNKGNSDPQALDKLLERLNDESTNKEK 483
DB 340 -LTTSEKLSASDGIYFNPK-DIVETATAYIVRG--DHFYIRK-----SNQIGQPT 389
QY 484 LVDDLAFAPITHEBRLKPNQIEYTEDEVRIQLADKYTTSQYTFDEHDIIISDEGD 543
DB 390 LPNNSLATPSP-SLPINQTSHEKHE-----EDQYGDARIIIAEDS 431
QY 544 AYVTPHMGSHWIGK 558
DB 432 GFVMSHGDHNYFFK 446

RESULT 10

US-09-536-784-182
Sequence 182, Application US/09536784
Patent No. 6573082
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/536,784
FILING DATE: 30-Oct-1997
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: OCT-30-1997
ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks
REGISTRATION NUMBER: PB340P3
REFERENCE/DOCKET NUMBER: PB340P3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 182:
SEQUENCE CHARACTERISTICS:
LENGTH: 447 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 182:
US-09-536-784-182

Query Match 28.5%; Score 1187.5; DB 4; Length 447;
Best Local Similarity 46.3%; Pred. No. 5.5e-84;
Matches 257; Conservative 72; Mismatches 117; Indels 109; Gaps 10;
QY 4 LGLVARTVKNRRSYIDGKQATKNTLTPDEVSKREGINAEOIVIKITDQGYVTSHG 63
DB 1 LNQHRSOENKDNRRSYVDGSSQSKSENLTPDQVSQKREGIOAEQIVIKITDQGYVTSHG 60
QY 64 DHYHYNGKVPYDALISELLMKDPNYKDKEDIVNEVGVIVKVDGKYVYLLDAHA 123

```

Db 61 DHYHYNKQVPPDALFSEELLMKDPNYOLKDAIYNEVGYYIIKVDGKYVYLLDAHA 120
Qy 124 DNVPRKEINFKQKSHQREGCTPRNGCAVALASQGRYTTDDYITNASHIIEDTDA 183
Db 121 DNVPRKEDINQKQKHVKNDE---KVSNAVAASQGRYTTNDYVNPADIIEDTENA 176
Qy 184 YIVPHGDHYHYIIPKNELSASELAAEAFLSGRGNLSNRTYRQNSDNTSRTNWPSVN 243
Db 177 YIVPHGHHYIIPKSDLSASELAAKAHLAGKNMPSQLSYSTASDN----- 224
Qy 244 PGTNTNTSNNSNTNSQASQSDIDSLKQLYKPLSQHVESDGLVDPDAQITRTARG 303
Db 225 ---NTQSAKSTSKPANKSENILQSLKELYDSPAQRYSESDGLVDPDAKIIISRTENG 280
Qy 304 VAVPHGDHYHYIPYQWSELEERIRIIPLRKSNHWVPDSRPQSPQPTPEPSPGP 363
Db 281 VAIPIGDHYHYIPYKSLALEEKIARMP----- 309
Qy 364 APNEKIDNSSIVSQLVKRVGEGYFEEKISRYVFAKDLPEETVKNLESKLQESVSH 423
Db 310 ---ISGTGSTVSTNAK-----PNEVYSSLSGLSSNPS--- 339
Qy 424 TLTKAKENVAPPDQSFYDKAYNLLTEAHKALFXNKGNSDFQALDKLERLNDSTNKEK 483
Db 340 -LTSKELSSASDGYIFNPK-DIVETATAYIVRHG--DHFIYIK-----SNQIGQPT 389
Qy 484 LVDLLAFLAPITHERIGKPKNSQLEYTEDEVRIQLADKYTTSQGYTFDEHDIISDGD 543
Db 390 LPNNSLAPSP-SLEINPGTSHKHE-----EDGYGDANRIIAEDES 431
Qy 544 AYVTPHMGSHWICK 558
Db 432 GFVMSHGDHNYFFK 446

```

```

RESULT 11
US-09-134-001C-5080
; Sequence 5080, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5080
; LENGTH: 3696
; TYPE: PR1
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5080

```

```

Query Match 3.9%; Score 163; DB 4; Length 3696;
Best Local Similarity 18.5%; Pred. No. 0.0054;
Matches 183; Conservative 141; Mismatches 348; Indels 316; Gaps 42;

```

```

Qy 15 NNRVSYI--DGKQATOKTENLTPDEVSKEGINAQIYIKITD-----QGYVTSHGHHY 67
Db 1299 NNQIDGIVSDRQSIINA---ITPDTSIRMAKN--DIDIKADKKIKIQRINDATDEBIQ 1353
Qy 68 YVNGKVPYDAIISEELMKDPYKLDKDIVNEVGKGVYIKV----- 110
Db 1354 EANKRIE-----EATIEAKDNQIRNSTDOVNEAKTNGIKNTENTPATYKSEARQAVQ 1408
Qy 111 ---GKYVYVLDAAHADNVRTKEEINR-----OKQESHQREGG--TPRNDGAVALA 157
Db 1409 NKAHQIINHIOHTPDATNEEKQEAIRNVSASIELARVQAOINAEHTTQGVYKTIKMDAITSL 1468

```

```

Qy 158 RSQGRYTTDDGYIFENASDIIEDTGDYIVPHGDHYHYIPKNELSASELAAEAFLSGRGN 217
Db 1469 R-----INAOVEKE-----SANA 1483
Qy 218 LNSRTRYRQ---NSDNTS-----RTNWPSVSNPGTTNTN--TSNNSNTNSQASQSDND 266
Db 1484 IEKATQOQTQFINNNDVATDEKEVANLVLATQKSLDININSLSNNDVENAKVAGINE 1543
Qy 267 IDSLKQLYKPLSQHVESDGLVDPDA-QITSTRARGVAPPHGDHYHYIPYQWSELEE 325
Db 1544 IANVL-----PATVAKSRAKDIQKLAQOINQIOTHTQATTEE 1582
Qy 326 RIARIIPLYRNSH-----WVDSREQSPQPTP--- 355
Db 1583 KEAIIQLANQKSNKARTALQNEHSNNGVQAQKSGIHIEILVMPDAHKKSAKQSIDKY 1642
Qy 356 -EPSPGPQADNLKIDNSSIVSQLV--VRKVGEGYFEEKISRYVFAKDLPESTVKNLE 412
Db 1643 NEQNTINTTDPADTEEEKOKALDKLIAKDGYNKVVDAQNTNQVSDAKTEAIDITNIQ 1702
Qy 413 SKLSKQESVSHITLAKKE-----NVAPROE-----FYDKAYNLL----- 447
Db 1703 ANVAKKPSARVELDSKEEDLKRQINATPNATEEKKODAIQRLNGKRDENVKLLINQDRDN 1762
Qy 448 -TEAHK-----ALFXNKGNSDFQALDKLERL-----NDESTNKEKLVDDL 489
Db 1763 EVEGHKINIGLOBLETIHANPRKSD--ALQSLQRTKISQTELLINNKQDATNEEK---DEA 1817
Qy 490 AFLAPITHERPL-----GKPSQLEYTEDE--VRIADL-----ADKYTSDGYIFDEH 535
Db 1818 KRLLEISKNKITITNINQQTNNQVDAKNQNMENIATIIIPATTIKTDAKTAIDKAEQOV 1877
Qy 536 DIISDEGDYVTPHMGSHWICKDSLDPKEVAAQAATKEKGLPSPSDAIVKANPTGDS 595
Db 1878 TTINGNDA-----TDEEAKERKLVKAKI-----EAKSITNSD 1913
Qy 596 AAAYNRVKGSEKRIPLVRLPYMVEHTEVVKGNLIIIPKHQYHNHIFAMPDHTYKAP-- 653
Db 1914 TE---REVNGAKTNGLEKINNIOPTSTQKTNAKEINDKAOEOLQINNTDABEEKQE 1970
Qy 654 -----NGYTLEDL-----FATIKYVHPDERP--HSNDGWNAS 686
Db 1971 ATNEVNAGLAQAIQINNNAHSTQVNESKTSIATIKSVQGNVLIKPTAINSLTQEANNO 2030
Qy 687 EHVGGKDHSEDPKNF-----KADE--EPVETRAEPVPOVETKVA-----Q 730
Db 2031 KTLGNGGNATDDEKAAKQVLTKLNEQIQIHESIQDQNDVNVKQAITAIKLINANA 2090
Qy 731 LKEAEVL-----LAKVTDSLKAN--AT-----ETLAGLRNLL-----TLQIMD 767
Db 2091 HKRODAINILNLNLESKSDIRANQDATTEEKNTALQSIDTTLQARNNINGANTNALVD 2150
Qy 768 NNSIWAERKLLALLKGSNPSSVSKEKI 795
Db 2151 EN-LEDGKQKQIRVLSTQTKTQAKADI 2177

```

```

RESULT 12
US-09-134-001C-3159
; Sequence 3159, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14

```

NUMBER OF SEQ ID NOS: 5674
 SEQ ID NO: 3159
 LENGTH: 10182
 TYPE: PRF
 ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-3159

Query Match 3.8%; Score 159; DB 4; Length 10182;
 Best Local Similarity 18.8%; Pred. No. 0.055;
 Matches 189; Conservative 126; Mismatches 368; Indels 320; Gaps 43;

```

13 KENNRSYIDGKATKENTLPEDESKKEGIN-----AEQIVIKITDGYVTSH 62
   :::::
5419 OKSGEELVNGSVNTRSEVE-----EHLNEKSLNNAMKQJLDKVAEKTNYK-QSDYIINDS 5473
   :::::
63 GGHYHYNGKV-PYDALISE-----ELMK-----DPNKLDEDEIVGVKGYVI 107
   :::::
5474 TEHQRGYDALOEAENINEIGNPTLNKSEIEQKLOQLTDAQNALQSHLEBAKNNAIT 5533
   :::::
108 KYDGKYYVYLKDAH--ADNVATKEEINRQKQESQHEGCTPRND-----GAVLARS 159
   :::::
5534 GIN--KLTLNDARQKAIENYQAQQTIRAVNQQLLDEHINTAMQALDKVGOQNNVHQ 5591
   :::::
160 QGRYTTDDGYIFENASDIIEDTDAVIVPHGDHYVYIPKREL--SASELAAEAFLSGRN 217
   :::::
5592 QSNYFNEDEQPKHYNNSVQAQGTIIDKLQD--PIMKKEIEQALINQINTTQALSGENK 5649
   :::::
218 L-----SNSRTYRQNSDNTSRNTWPSVSNPCTTNTNTSNNSTNSQASQSDNDLSLK 272
   :::::
5650 LHTDESTNRQIEGLSSLTNAQINAEKDLVNOAKTIDVA-----QKLAAKEINSAMS 5703
   :::::
273 QLYKPLSORHVESDGLVDFDPAQITSRARGVAVPHGDHYHFIPIYQSGMSELEERIRIIP 332
   :::::
5704 NL-----RDI-----QNKEDIRSSAYINADPRKATAYQALQNAENITINATP 5747
   :::::
333 LKRSNHWVPSRPEQSPQPTPEPSPGQAPALNKLIDSNSSLVSQLVRKVGGEVFEK 392
   :::::
5748 -----NVELNKATIEQALSRVQAQ--DLD 5771
   :::::
393 GISRYFAKDLPSSETVKNLES-----KLSQESVSHLTAK 428
   :::::
5772 GVQQLANNAQOATQYVNGNLNSLDGQKRELINLLINSANTRTKVQEBELNATLETNHAEMEL 5831
   :::::
429 KENVAFRD-----QEFYKAYNLLTEAHKALFYKGRNSDPQALDKLBER 473
   :::::
5832 RNSVQVNDVQKQSNVYNEDQEPQHNYDNAVN--EAQATINNAPVLDKLAIEHLLTQ 5888
   :::::
474 LN--DESTNKEKLVDDLAFIAPITHPERLGPNSQIEY-----TEDEVY-IAQ-- 519
   :::::
5889 VNTTKALHGAQKLTQDQAAETGIRGLSINEPQKNAEYAKYTAATTDEVANINQEA 5948
   :::::
520 -----LADKYTT--SDGYTFDEHD-----IISDEGDAVYTFHMGH 552
   :::::
5949 TLDTAMLGARKSKIKDNNDTKNSKSYINEBDQOQAAVNNVNAQVVIDEQATLS----- 6003
   :::::
553 SHWIGKDSLDEKVAQAAYTEKEGILPSPRADYKANPTGSA-----AIVYRVAGGKR 608
   :::::
6004 -----SDTINQLANAVTQAKSNL--HGDTKLQDKQSAKQTTIQOLQNLNSAQKH 6050
   :::::
609 -----IPVRLPYMVEH-----TVEYKQGNLI--IPKHDYH- 638
   :::::
6051 MEDSLIDNSTRQVQHDLTQAQALDGLMGALKESTKDYTNIVSNQNYINAEBSKQAYD 6110
   :::::
639 -----NIFKMFDDHYT-KAPN-GYTLDEL--- 661
   :::::
6111 AAVQNAQNIINGTNOPTINKGNVTTATQTYKNTKDLGDGHRLEBAKNNANQITRMJSLN 6170
   :::::
662 -----FATIKYVEHPDERPHSNDGWNASEHVLGKKDH-----SEDP 699
   :::::
6171 NNAQKAEKLVNSASTLEVOQNLOTQAQDLNAMBELROSI-AKQDQYKADSKYINEDB 6229
   :::::
700 NKNFKADE-----EPVEETPAEPEVQVETEKYEAQLKEAEVLL--AKYTDSSLKANATE 752
   :::::

```

DB 6230 QIKONYDAVQREVTINETONPELLKANIDATOSVQNAEQALHGAERKLNQDKTQSTTE 6289
 QY 753 TLAGLRNNLTQIMONNSIMAEAKLLALLKSGNSPSSVEKRI 795
 DB 6290 -LDGL-TDITL-----DAQREKLRQEQINTNSKRDIDKQKI 6321

RESULT 13
 US-09-513-783A-6
 Sequence 6, Application US/09513783A
 Patent No. 6416959
 GENERAL INFORMATION:
 APPLICANT: Giuliano, Kenneth A.
 APPLICANT: Kapur, Ravi
 TITLE OF INVENTION: A System for Cell Based Screening
 FILE REFERENCE: 97-022-LI
 CURRENT APPLICATION NUMBER: US/09/513,783A
 CURRENT FILING DATE: 2000-02-25
 NUMBER OF SEQ ID NOS: 180
 SOFTWARE: Patent In Ver. 2.0
 SEQ ID NO 6
 LENGTH: 812
 TYPE: PRF
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence:
 OTHER INFORMATION: EYFP-DEAD-MARKDM construct
 US-09-513-783A-6

Query Match 3.5%; Score 147; DB 4; Length 812;
 Best Local Similarity 19.8%; Pred. No. 0.0086;
 Matches 175; Conservative 103; Mismatches 331; Indels 274; Gaps 39;

```

53 ITDQY-----VTSQGHYHYNGKVVPYDAIISEELMKDP-NYKLDK-----DIY 98
   :::::
62 VTFPGGLQCFARYDPDHMKQHFPSKAMEGYOERTTIFPKDDGNYKTRAEVYFEGDTIV 121
   :::::
99 NEVK-GGYVTKVDG-----KYVVYLKDAHAANVATKEEINRQKQESHS-----Q 141
   :::::
122 NRIELKQIDFKEDGNILGHKLEYNNNSHNVYIMADKQKNGIKVFKIRNIEDSGVOLAD 181
   :::::
142 HREGGTPRNDGAVALARSGQRYTTDDGYIFENASDIIEDTD-----AYIVPHG 189
   :::::
182 HYQNTPIPIDGVLL-----PDNHYLSYQSALSKDPNEKRDMHVVILLEFYTAAGITIG 233
   :::::
190 DHYHYTPKKELSASEL-----AAABA----- 210
   :::::
234 MDLKYPRDEADSDLSLVDALTEPPPEIEGIRKDFMALAEAPYDDIVGETVEKTEFI 293
   :::::
211 -FLSGRGNLSNRTYRQNSDNTSRNTWPSVSNPCTTNTNTSNNSTNSQASQSDIDS 269
   :::::
294 PLDDGDEKTNSESKKPCPLD-TSQVEGIPS-SKP-TLIANDHGEGNNTAGSPD--- 347
   :::::
270 LKQLYKPLP--LSORHVESDGLVDFDPAQITSRARGVAVPHGDHYHFIPIYQSGMSELEER 326
   :::::
348 FLEERVDYDYQSSQWPEFASFCFQPOQVLDTQ--AEFPNEH-----RDGQ 393
   :::::
327 IARIITPLKRSNHWVPSRPEQSPQPTPEPSPGQAPALNKLIDSNSSLVSQLVRKVGES 386
   :::::
394 LADLLFVSSGPTNASAFTERDNPSS--EDSYGMIPCSF--ASTAVASQEWISGADN 445
   :::::
387 YVFEKGISRYFAKDLPSSETVKNLE-----SKLSQESVSHLTAKKENVAVRDQEFVD 441
   :::::
446 SPQSCSCVS-----PEVTTETLOPATELSKAAEVESVKEQLPKK----- 484
   :::::
442 KAVNLTTEAHKALFYKNGRNSDPQALDKLBERLND-----ESTNKEKLVDDLAFIAPITH 497
   :::::
485 -----ALETMAEQTTDVVHSPSTDTTGPRTAALANDIE- 519
   :::::
498 PERLGPNSQIETDEDEVYIAQLADKYTTSDGYTFDEHDIISDEGDAY----- 545
   :::::
520 -----BITRPDVILANVTQPTESDWFIAQDMELLTGTETAHANNTIILPTEPDE 568
   :::::

```

QY 546 -----VTPHMGSHWIGKDSLSDEKEVA-----QAYTEKGIUPEPDADVK 588
 Db 569 SSTKVAAPMEBEIIPGNDTTSPKETETTLPIKMDLAPEDVLKRETELAP-----AGM 624
 QY 569 ANPTGDSAAAIYNYKGEKRIPLVRLPYWVEHTVEYKGNLIIPIKHVHNKIFAMFDH 648
 Db 625 VLSLEIBALAKNDVRSAB-IPVAO-ETVASETEVLALTEVLP----- 666
 QY 649 TYKANGTLEBDFATIKYVEHP--DERPHSNDGMN-ASPHVAGKKO-HSEDPNKNFK 704
 Db 667 -----SDPITTLTKDVTLPLEAERPLVDTMTPSLETMTGKETAPETETNLGMA 716
 QY 705 ADEEVEETPAE-----PEVPOVETEKVAOLKEAEVLLAKVTDSSLKANATETLAG 756
 Db 717 KDMSPLEPSEBVTLGKDVILPETKVAEFNNV--TPIJSEEV--TSVKDMSPSA---ETEAR 770
 QY 757 LRNNITLQ-----INDNNSIMAEKLLALKGNSSSVSKK 794
 Db 771 LAKNADLHSGTELIYDVS--MAPASDLALPLETKVATVPIDK 811

RESULT 14
 5231168-2
 ; Patent No. 5231168
 ; APPLICANT: DZIEGIEL, MORTEN, BORRE, MARTIN, JEPSEN, SOREN;
 ; VUUST, JENS, RIENECK, KLAUS, WIND, ANNETTE, JAKOBSEN, PALLE H.
 ; TITLE OF INVENTION: MALARIA ANTIGEN
 ; NUMBER OF SEQUENCES: 19
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/409,658
 ; FILING DATE: 18-SEP-1989
 ; SEQ ID NO:2:
 ; LENGTH: 783
 5231168-2

Query Match 3.5%; Score 146; DB 6; Length 783;
 Best Local Similarity 20.2%; Pred. No. 0.0097;
 Matches 141; Conservative 106; Mismatches 278; Indels 174; Gaps 35;

QY 178 EDTGAVIYPHGDHYHYPKIELSASELAAPFISGKGNLSNSTYRQ--NSNTSRT 235
 Db 35 EESGSSKLV-----DNDEGFEEKHENFSS--EVSNSLENENFEVSDK---- 77
 QY 236 NWVPSVPGTNTNTSNTSNTSOASNDISLLKOLYKLSORHVS--DGLV--- 290
 Db 78 ----SVTPAEHEEVSESNPEAPNESSSIEAHQ--EIVPQNDSESGSLVNE 131
 QY 291 ----FDPAQITS-----RTARGVAVPHGDHYHFIYSQMSLEERIRARI 331
 Db 132 EGDPEPNHBEPEPQNDSELSNELVSEKSVSEPR-AEHVEIVSEKSVSEPAHV-EIV 189
 QY 332 PLRYNSNMWPDSPREQSPQPTPEPS--PGQAPNL----KIDSSSLVSQLVKRG 384
 Db 190 SEKSTSE--PAEHVESVSEOSNNEPSEKQDPVSKPFEIEIEKDVOPKIVD--LQITE 244
 QY 385 EGYVF-----BEKGISRYVFAKDLPSSETVY--NLESKLSKQESVHTLTAKK-----EN 431
 Db 245 PNFUDSQNPQEPVPSFVKIEKVSSENKASVDPVYKEKENVSEVVEEKNQSGESIBE 304
 QY 432 VAPRDOEFYDKAYNLLTEAHKA-----LFXNKGKNSDFALDKLERLND 477
 Db 305 I PVNEDEFEDVYTEOLDLHKTVDPEIVEEIPSELHENEVAHDEIVEIEVEFPEPNQ 364
 QY 478 STNNEKLVDDLAFILA-PIYTHEBRLGKRSQIEYEDE-VRIAQDLADKTTSDGTFIPEH 535
 Db 365 NEFOEINDDXSANIQHEIVEVEEILPEBDKNEKEHEIVEVEEILPEBDKNEKG----QH 420
 QY 536 DIISGEGAYVYTPHMGSHWIGKDSLSDEKEVAADYATYKKEGIIPLPSPADVANKAPTGS 595
 Db 421 EIVEVE---EILPE-----DDNKEVHEIVEVEEILP--BEKKEKGQHEIVE 463
 QY 596 AAALYNYKGEKRIPLVRLPYWVEH-TVEVKGNLIIPIKHVHNKIFAMFDHTYKAPN 654

Db 464 VEIILPEBDKNEK-----VEHEIVEVEE--ILPEDKNEKG-----QHEIVEVE 503
 QY 655 GYTLEDDLPATIKYVVEHDE-RPHSNDGMNAA-----EHVLGKKDSE----- 697
 Db 504 EILPEDKNEKQHEIVEVEEILPEDKNEKGQHEIVEVEEILPEDKNEKGQHEIVEVEI 563
 QY 698 -DPNKNFKADEPVETPAPEVPOVETEKVAOLKEAEVLLAKVTDSSLKANATETLAG 756
 Db 564 LPEDKNEKQHEIVEVEEILPEDKNEKQHEIVEVEEILPEIVE-----IEVPSQ 614
 QY 757 LRNNITLQIMNNSIMAEKLLALKGNSSSVSKK 795
 Db 615 TNNNENIE-----TIKPEKK-----NEFSVEEKAI 640

RESULT 15
 US-09-541-782-6
 ; Sequence 6, Application US/09541782
 ; Patent No. 6284480
 ; GENERAL INFORMATION:
 ; APPLICANT: Nislow, Corey
 ; APPLICANT: Sakowicz, Roman
 ; APPLICANT: Berand, Christophe
 ; TITLE OF INVENTION: Antifungal Assay
 ; FILE REFERENCE: 1015
 ; CURRENT APPLICATION NUMBER: US/09/541,782
 ; CURRENT FILING DATE: 2000-04-03
 ; NUMBER OF SEQ ID NOS: 10
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 6
 ; LENGTH: 1073
 ; TYPE: PR1
 ; ORGANISM: S.pombe
 US-09-541-782-6

Query Match 3.5%; Score 145.5; DB 3; Length 1073;
 Best Local Similarity 18.6%; Pred. No. 0.017;
 Matches 182; Conservative 136; Mismatches 357; Indels 301; Gaps 41;

QY 23 GKQATQKTEMLTPREVSREGINE-----QIVIKI--TDQGVYTHSGHYHYNN 70
 Db 147 GGTGKTYTWSGD-LSDSDSLSEAGLIPRALYQLFSSLDNSQNEYAVK-CSTYELYN 204
 QY 71 GKVPYDALISEELMKDPNYLTKDED-----IVNEVGKGYIKYDGYVYLKDAAHADN 125
 Db 205 EET-RDLVSEEL--RKPARVFDTSRGNVVIIGIESY--IKAGDGLRLRREGSHRQ 260
 QY 126 VRTKEEINRQKQHS-----QHREGTFRNDGAVALARSGRYTTDDGYIFNASDIIDT 180
 Db 261 VAATKCNDSLRSRSHIFITLHARKVSGMTDETSLTINN-----NSDDLRL-- 307
 QY 181 GDAIVPHGDHYHYPKIELSASELAAPFISGKGNLSNSTYRQ----- 223
 Db 308 ----ASKLHMVLDAGENI--GRSGAENRABETGMINOSLTLGCVI 349
 QY 224 ----YR-----RONSMTSRTNWPVSYNPQTNTNTSNTSNTSOASQS 264
 Db 350 NALVEKAMHIYRRESKULRLLQDLSGKTKTSMIVTYS--STNTNLEETISTLEVARA 406
 QY 265 NDIDSLKOLYKLSORHVESDGLVPDPAOI-----TSRTARGVAVPHGDHYHFIYSQ 319
 Db 407 K---SIRKPPQNNOLVFERKVLIKDLVLDIEHLKNDLNAIRKKNQVYLAESTYKEIMBRVQ 463
 QY 320 MSEL-----EEIATIIPLRYNSNMWPDSPREQSPQPTPEPSGQPOAPNLKI----- 369
 Db 464 NKDILCOEQARKLEVLDLNIVS-----SRQOLOVYSKSNQEHKEVEALQLOLQVNSSTE 517
 QY 370 ----DSNSLVSQLVKRGVGEYFEEKGISRYVFAKDLPSSETVKNLESKLSKQESVSH 424
 Db 518 LESVKSNEKIKNELVLEIEKRYETNEAKITTVAVDLSQYRESKEYIASLYEKDRT 577
 QY 425 LTAKKENVAPRDOEFYDKAYNLL-----TEAH 451

```

Db      578  ERNNKEN-----ENNPMNLKFNLLTWLRBSFHSGFTDETNGVFTLLDNFNASMBELLNTHSN 633
Qy      452  KALPFNKRNRGNDFOALDKLIER-----INDESTNKEKIVDDLALFLAFLAYT 496
Db      634  QLLISMWTIKTEHPQSLDBALQSAARSSCAVPNSSLDLIVSELKDSKNSLLDLAHLSDLOIS 693
Qy      497  -HPERLQKPN-----QIEYTEDE-----VR--IAQL 520
Db      694  MSSQQLGNGISSLLELQKDMKESRYOLVOELRSIYNQIHTEBSQKMLMGVRNDIDAL 753
Qy      521  ADKTYTS--DGYIFEDHDIISDEGDAYVTPHMGSHWIKGKOSLSDEKYAAQAYTKEGKI 578
Db      754  VKTCSTSLNDADI--LSDYISIDQSKSPESKOODLIANIGKIVSNPLQCONESLYLK----- 808
Qy      579  LPSPDADVKNPCTDSAAAI--YVRVGEKRIPIVR-----LPVWVEHTVEVK 625
Db      809  -----ADILSHLNDTNSNIRKANEIMNNSSEETLRMAAGOAELVGNANKERIKQTV-- 860
Qy      626  NGNLIIPKHD-YHNIKFAMPDDHTYKAPNG-----YTLLEDIFATIKYVEHPDERP 676
Db      861  NGSQQLDSKSKAIAHSNRSRMYDHCIALAESQKQVLELVQTLDRLLQKVK-----E 911
Qy      677  HSNQGWGASHVILGKXDHSEDPNKNFPADEBPVEBETPAEPVPEVTEENKXAQLKEAV 736
Db      912  HSEDYTKKHQOQLDLLESLVGANNDL-----IDSIKTPHTELQ----- 950
Qy      737  LLAKVTSLSLK-----ANATEFLAGLRN-----NLTLQIMDNNSIMAAE-----EKL 780
Db      951  ---KITDHVLKGTSTLANHTNELLGQBSCLNLETTIEDISLVLEKTTGTTPSKRELPA 1007
Qy      781  LKGSNPSVSKEKIN 796
Db      1008  TPSWTRDSSLIKETTN 1023

RESULT 16
US-09-723-820-6
: Sequence 6, Application US/09723820
: Patent No. 6468760
: GENERAL INFORMATION:
: APPLICANT: Nislow, Corey
: APPLICANT: Sakowicz, Roman
: APPLICANT: Beraud, Christophe
: TITLE OF INVENTION: Antifungal Assay
: FILE REFERENCE: 1015
: CURRENT APPLICATION NUMBER: US/09/723,820
: CURRENT FILING DATE: 2000-11-28
: PRIOR APPLICATION NUMBER: 09/541,782
: PRIOR FILING DATE: 2000-04-03
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 6
: LENGTH: 1073
: TYPE: PRT
: ORGANISM: S.pombe
US-09-723-820-6

```

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/961,083
 FILING DATE:
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Brookes, A. Anders
 REGISTRATION NUMBER: 36,373
 REFERENCE/DOCKET NUMBER: PB34022
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 118:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1040 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-961-083-118

Query Match: 3.5%; Score 144; DB 3; Length 1040;
 Best Local Similarity 19.2%; Pred. No. 0.02;
 Matches 154; Conservative 111; Mismatches 284; Indels 252; Gaps 38;

107 IKVDGKYV-YLKDAADNVRTKEEINRQKQEHSGHREGTFRND-----G 152
 76 LKIEGYOYIGYK-TKKQDNTLSRTVDGKYSAGDSQPNSTKTSADVHSADLEWNOGOG 134
 153 AVALARSGRYTDDGYIFNASDIIEDTGDAVYVPHGDHYIIPKNELSASELAAREFL 212
 135 KVSU---QGEASGDDG-----LSEKSIADNLSNDSPA 166
 213 SGRGLNSRTYRRQNSDNTSRTNVPSVSNPGTTNTNTSNNSTNSQASQNDIDSLK 272
 167 S-----QVEQNPDHKGESVVRPTVPEQGNFVSATVQSAREEVLATND-----R 211
 273 QLYKPLSGRHVESG-----LVF-----DPAQITS 298
 212 PEYKLPLETGTQGEHGEAAVRBDLPYTKPLTKGTQGEHGEAAVREERAYTEP 271
 299 RTARCAVAVPHGDHYIIPYSQWSELEERARIIPRYSNMHWVPSRPEQSPQPTPEPS 358
 272 LATKGTQGEHGE-----GKATVEETLEYTEPVAATGT-----QPEHGEERAVEEL 319
 359 PGPOPA-----PML-----KIDNSSLVSQLVR---KVGEGYFEE 391
 320 PALEVTNRTETIONIPTYTTEIIOPTLLKNRKRTEROGAQTRTIOYEDYIVNGNVET 379
 392 KGISRYVAFKOLPSTVKNLSKLSKQSVSHTLAKKENVAPRQGEFYDAVNL--TE 449
 380 KEVSTTEVA---PNEVVAVGVGLVKVKTVEITNLTKVEN---KKSITVSYNLTDTTS 431
 450 AHKALFXNKGNSDFQALDKLIERLNDESTNKEKLVLDLAFIAP---TH-PERIGKPN 505
 432 AYVS-----AKTOVPHG-DKLIVKEVDINPAKEQVYISG-LDYTTTYTKHITVYLGNN 484
 506 SQIEYTEDEVARIAQLADRYTSDGYTFEBHDIIISQEGDAVYTPHMGSHMWIGKDSLSKE 565
 485 E-ENTETSTDFOL--EYKIKIEDIDISVELYGENRNY-----RRYL---SLSEAP 530
 566 KVAAGAYRKEGKILPSPDOADVKANPTGSAAIYNRVKGKRIPLVVLPTWVEHT--- 621
 531 TDTAYFYKVK-----SDRFKEMTLPVKS-----ITENTDGT 563
 622 -VEVNGNLIIPKDHYNIKFAMPDDHTY-----KAPNGYT-LEDLFAITK-----Y 667

DB 564 KTVAVADQVLEEGTDGYK-----DDYFTVAKSKAQOPGVYTSFKQLYTAMQSNLSGV 616
 QY 668 YVEHDEPHNDGNGASERVLGKKHS-----EDPRKNKABE--EPVEET 713
 DB 617 YTLASDM-----TADSVSLDGQTSYLTGAFTSLGSDGTSYAIYDLKKPLPDT 667
 QY 714 PAEPVPOVETEKVEAQLKEAEVLIKAVTDSLSL-----KANATEFLAGRNNLTQI 765
 DB 668 LMGATVRLDITKYTSADSKENVALAKAANSANNINNAVEGKISGAKSVAGIVASATV 727
 QY 766 MDNNSIMAEKLLALLKGSN 786
 DB 728 IENSFTG---KLIAHODSN 745

RESULT 18
 US-09-536-784-118
 Sequence 118, Application US/09536784
 Patent No. 6573082
 GENERAL INFORMATION:

APPLICANT: Choi et. al.
 TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
 NUMBER OF SEQUENCES: 452
 CORRESPONDENCE ADDRESS:
 ADDRESS: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/536,784
 FILING DATE: 30-Oct-1997
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/961,083
 FILING DATE: OCT-30-1997

ATTORNEY/AGENT INFORMATION:

NAME: Michelle S. Marks
 REGISTRATION NUMBER: 41,971

TELECOMMUNICATION INFORMATION:

REFERENCE/DOCKET NUMBER: PB340P3
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 118:

SEQUENCE CHARACTERISTICS:

LENGTH: 1040 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 118:

US-09-536-784-118

Query Match: 3.5%; Score 144; DB 4; Length 1040;
 Best Local Similarity 19.2%; Pred. No. 0.02;
 Matches 154; Conservative 111; Mismatches 284; Indels 252; Gaps 38;

107 IKVDGKYV-YLKDAADNVRTKEEINRQKQEHSGHREGTFRND-----G 152
 76 LKIEGYOYIGYK-TKKQDNTLSRTVDGKYSAGDSQPNSTKTSADVHSADLEWNOGOG 134
 153 AVALARSGRYTDDGYIFNASDIIEDTGDAVYVPHGDHYIIPKNELSASELAAREFL 212
 135 KVSU---QGEASGDDG-----LSEKSIADNLSNDSPA 166
 213 SGRGLNSRTYRRQNSDNTSRTNVPSVSNPGTTNTNTSNNSTNSQASQNDIDSLK 272

Db 167 S-----OVEDNDHKGESVVRPIVEQGNPVASATTVOGSAEEVLAATTND-----R 211
Qy 273 QLYKLPISGRHVESDQ-----LVF-----DPAQITS 298
Db 212 PEYKLPLETKGTQOEPOHGEBAVREDLPYTKPLETKGTQOGHEBAVREEPYTP 271
Qy 299 RTARGAVPHGDHYHPIYSQMSLEERIAIIPLRYSNNHWVDPSPQSPQPTPEPS 358
Db 272 LATKGTQOERGHE-----GKATVEETLETPYATKGT-----QEPHEGERXVEEL 319
Qy 359 PEPQPA-----PNL-----KIDSNSLVSQLVR---KXGEGVFE 391
Db 320 PALEVTTRNRTIONIPTYTEEIODETLKNRKRKIEROGAGTRITQYEDYIVNGVAVPT 379
Qy 392 KGISRYVFAPKDPSETVKNLESKLSKQESVSHTLTKAKENVAIPRODEFKAYNLL--TE 449
Db 380 KEVSTTEVA---PVNEVAVVGLVKKPVEITNLTKVN---KKSITVSYNLDITS 431
Qy 450 AAKALFXNKGNSDFQALDKLERLNDESTNKEKLVDLLAFAPL---TH--PERLGEN 505
Db 432 AVVS-----AKTQVFHG--DKLVKAVDIEHPAKEQVIG--LDYTPYTKHLYNNGEN 484
Qy 506 SQIETTEDEVRIQALADKTTSDGYTFDEHDIISDEGDAYVTPHMGSHWIGKDISDKE 565
Db 485 E--EMETSTJODPOL--EYKKLEIKDIDVVELYGENDEY-----RRYL--SLSEAP 530
Qy 566 KVAQAAYTKEKGIILPPSPADVAVKANPTGSAAIYRVKGEKRIPLVRLPVVVEHT--- 621
Db 531 TDTAKYFVYVK-----SRPEKMTLPVVS-----ITENTDGY 563
Qy 622 -VEVKNGLIIPKHNYHNIRKAMPDDHTY-----KAPNGYT--LEDLFATIK---Y 667
Db 564 KVTVAVDQJVEBGTDGYK-----DDYFTVAKSKAEQPGYTTSKQLVTMOSLSCV 616
Qy 668 YVEHDERPHSDMGNSAEHVLGKKDHS-----BDPNKNEPADE--EPVEET 713
Db 617 YLIAADM-----TADSVLSGDKQTSYLTGAFTGSLIGSDGTSAIYDLKKPLPDT 667
Qy 714 PAEPVPOVETEKEVAQLKEAEVLLAKVTDSSL-----KANATETLAGRNMLTQI 765
Db 668 LNGATVRDIDITVSADSKENVAALKAANSANNINNAVEGKISGAKSVAGVASKATNTV 727
Qy 766 MDNNSIMAEKLLALLKGSN 786
Db 728 IENSFTG---KLINHQDSN 745

RESULT 19
US-09-816-703A-2
; Sequence 2, Application US/09816703A
; Patent No. 6455026
; GENERAL INFORMATION:
; APPLICANT: Melcher, Thorsten
; APPLICANT: Mueller, Sabine
; APPLICANT: Chin, Daniel
; TITLE OF INVENTION: Use of Protein Tyrosine Phosphatase Zeta as a Biomolecular Target
; TITLE OF INVENTION: Treatment and Visualization of Brain Tumors
; FILE REFERENCE: 262/235 AG
; CURRENT APPLICATION NUMBER: US/09/816, 703A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 2314
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: Reference
; LOCATION: (1)..(2314)
; OTHER INFORMATION: Krueger, N.X. and Saito, H.: A human transmembrane protein-tyrosi
; OTHER INFORMATION: ne-phosphatase, PTP zeta, is expressed in brain and has an N-term
; OTHER INFORMATION: inal receptor domain homologous to carbonic anhydrases; Proc. Nat
; OTHER INFORMATION: l. Acad. Sci. USA 89 (16), 7417-7421 (1992

NAME/KEY: Reference
LOCATION: (1)..(2314)
OTHER INFORMATION: Levy, J.B., et al.: The cloning of a receptor-type protein tyrosi
OTHER INFORMATION: ne phosphatase expressed in the central nervous system; J. Biol.
NAME/KEY: SIGNAL
LOCATION: (1)..(24)
OTHER INFORMATION: By similarity
NAME/KEY: DOMAIN
LOCATION: (25)..(1635)
OTHER INFORMATION: Extracellular (potential)
NAME/KEY: misc_feature
LOCATION: (25)..(2314)
OTHER INFORMATION: Mature chain; protein-tyrosine phosphatase zeta
NAME/KEY: DOMAIN
LOCATION: (34)..(302)
OTHER INFORMATION: Carbonic-anhydrase like
NAME/KEY: SITE
LOCATION: (105)..(105)
OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
NAME/KEY: SITE
LOCATION: (134)..(134)
OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
NAME/KEY: SITE
LOCATION: (223)..(223)
OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
NAME/KEY: SITE
LOCATION: (232)..(232)
OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
NAME/KEY: DOMAIN
LOCATION: (312)..(406)
OTHER INFORMATION: Fibronectin Type-III
NAME/KEY: SITE
LOCATION: (324)..(324)
OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
NAME/KEY: SITE
LOCATION: (381)
OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
NAME/KEY: SITE
LOCATION: (497)..(497)
OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
NAME/KEY: SITE
LOCATION: (501)..(501)
OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
NAME/KEY: SITE
LOCATION: (552)..(552)
OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
NAME/KEY: BINDING
LOCATION: (587)..(587)
OTHER INFORMATION: Chondroitin sulfate (potential)
NAME/KEY: SITE
LOCATION: (602)..(602)
OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
NAME/KEY: SITE
LOCATION: (629)
OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
NAME/KEY: BINDING
LOCATION: (637)..(637)
OTHER INFORMATION: Chondroitin sulfate (potential)
NAME/KEY: SITE
LOCATION: (677)
OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
NAME/KEY: VARSPLIC
LOCATION: (755)..(1614)
OTHER INFORMATION: Splicing variant; missing (in short isoform)
NAME/KEY: BINDING
LOCATION: (997)..(997)
OTHER INFORMATION: Chondroitin sulfate (potential)
NAME/KEY: SITE
LOCATION: (1017)..(1017)
OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
NAME/KEY: SITE
LOCATION: (1050)..(1050)

```

; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1082)..(1082)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1122)..(1122)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1456)..(1456)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: BINDING
; LOCATION: (1548)..(1548)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: BINDING
; LOCATION: (1550)..(1550)
; OTHER INFORMATION: Chondroitin sulfate (potential)
; NAME/KEY: SITE
; LOCATION: (1561)..(1561)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: SITE
; LOCATION: (1617)..(1617)
; OTHER INFORMATION: Glycosylation; N-linked (GLCNAC...) (potential)
; NAME/KEY: TRANSMEM
; LOCATION: (1636)..(1661)
; OTHER INFORMATION: Transmembrane region; potential
; NAME/KEY: DOMAIN
; LOCATION: (1662)..(2314)
; OTHER INFORMATION: Cytoplasmic (potential)
; NAME/KEY: CONFLICT
; LOCATION: (1722)..(1728)
; OTHER INFORMATION: Missing (in ref. 2)
; NAME/KEY: DOMAIN
; LOCATION: (1744)..(1997)
; OTHER INFORMATION: Protein-tyrosine phosphatase
; NAME/KEY: ACT_SITE
; LOCATION: (1932)..(1932)
; OTHER INFORMATION: Active site; by similarity
; NAME/KEY: DOMAIN
; LOCATION: (1998)..(2314)
; OTHER INFORMATION: Protein-tyrosine phosphatase
; NAME/KEY: misc_feature
; LOCATION: (2222)..(2222)
; OTHER INFORMATION: Ancestral active site
US-09-816-703A-2

Query Match      3.4%; Score 141.5; DB 4; Length 2314;
Best Local Similarity 19.2%; Pred. No. 0.12;
Matches 169; Conservative 113; Mismatches 267; Indels 331; Gaps 43;

QY 51 IKITGGVVTSHGDHYHYNGKVPYDAIISELLMKDPYKXKDEDIVNEVGVIKVD 110
DB 964 VGVTYQSSLFSGPSHI-----PIKSSLIPTTASLIQPHALSG-----D 993
QY 111 GKYVYVLLKDAHADNVRTKEINROKOHSHQREGGTPRNDGVAL-----ARSGRYT 164
DB 994 GEM-----SGASSDS-----EFLLEDPTDGLTALNTSSPVSVAEFTYT 1030
QY 165 T----DDGTFNAGSIIEDTGDAYIVPHGDHYH-----IPK-----NELSASGLAAE 209
DB 1031 TSVFEDDKKALKSKSEIIVGNTELOIPSFENWVYSESTVPMWMDNVNKLNASIQETSV 1090
QY 210 AFLSGR-----NLSSRSYTR-RONSDDNTS-----RTNVVP 239
DB 1091 SISSTKGMPPGSLAHTTKVPDPHEISQVPENNFVSQOPHTVVSQASGDTSLKPVLSANSEP 1150
QY 240 SVSNPGIT-----NTNTSNNSVTNSQAS-QSNDIDSLKQKYKPLSQRHVES 286
DB 1151 ASSDASSEMSPSTQLLFYEISASFSFTEVLLQPFQASDVDTLKTIV--LPA-----VPS 1204
QY 287 DGLVDDPAQ-----ITSRTAKGVAVPHGDHYHPIFYQSQSELEERTIARIIPRY 335
DB 1205 DPILVETPKVDKISSTMLHLIVSNSASSEMMLHSTSVFQVDSPTSHMHSASLIQGLTISY 1264

```

```

QY 336 RSNHVPDSRREPQSPPTPEPSPGQAPAPLKLIDSNSLVSQLVKVGEGVEEKGIS 395
DB 1265 ASEKYEPLULKSSESHQVVP-----SLYSNDELFPQTANLEINQAH--PPKG-- 1308
QY 396 RYVPAKDLPS--EIVKULESKLSKQESVSHTLTKAKENAVR-----DOEFTYKAY 444
DB 1309 RHVFATPVLSDIEPLNTLINTKLHSDEI--LTSTKSVGCKVPAPIPTVASDPTFVSTDH 1365
QY 445 NL-LTEAH-----KALFXNKG----- 459
DB 1366 SVPIGNHVAITAVSPHRDGSVTSKLLFPESKATSELSHSAKSDAGLVGGEDEGDTDDG 1425
QY 460 -----RNSDFQALDKLERLNDSESTNKRKYVDLPLAPITHPERLGPNSQIEY---- 510
DB 1426 DDDDDRRSDGSLHKCMS--CSYRESQEKVND-----SDTHNSLMDQNNPISYLSSE 1478
QY 511 -TEDEVRIQLADKYTT-----SDGYTFDEHD1-----ISDEG 542
DB 1479 NSEEDNVTSVSSSQGDMRSPGKSPSANGLSOKHNDGK--EENDIQTGSAALLPLSPES 1536
QY 543 DAYTPHMGHSHWIGK---DSLSDKEKYA--AQAYTEK---GIL-----PSP 583
DB 1537 KAMAVLTSDEBSGGQGTSDSLNENETSDPSFADTNEKADGILLAGDSITPGFPSP 1596
QY 584 DADV-----KANPTGDSAAAIYVRKGEKR--IPLVRLPYWEHTVENVKNG 627
DB 1597 TSSVTSENSEVFHYVSEASNSHSHSRIGLAEGLESEKKAIVPIVYSALIFLCLVVLVG 1656
QY 628 NL-----IIPKHQYHNKIKAWPDHT--YKAPNGY 656
DB 1657 ILIYWRKCFQTAHFYLEDDSTSPRVISTPPTPIFISDVGVALPIKHFPKHAVADLHASSGF 1716
QY 657 TLEDLPATIK-YVYE-----HPERRPHSN 679
DB 1717 TEE--FETLKEFYQEVOSCTVDLGITADSSNHPDNK-HKN 1753

```

```

RESULT 20
US-07-920-281C-2
; Sequence 2, Application US/07920281C
; Patent No. 5739026
; GENERAL INFORMATION:
; APPLICANT: Garoff, Henrik
; APPLICANT: Liljestrom, Peter
; TITLE OF INVENTION: DNA Expression Systems Based on
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/920,281C
; FILING DATE: 13-AUG-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M.
; REGISTRATION NUMBER: 26,977
; REFERENCE/DOCKET NUMBER: 828-103P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-241-1300
; TELEFAX: 703-241-2848
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

```

LENGTH: 2431 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-07-920-281C-2

Query Match 3.4%; Score 141; DB 1; Length 2431;
 Best Local Similarity 19.7%; Pred. No. 0.14;
 Matches 137; Conservative 93; Mismatches 253; Indels 212; Gaps 34;

QY 56 QGVYVSHGHYHYNG-KVPYDAIISELLMKDPNKLKDEDI VNEVGKGYIKVDGKYY 114
 DB 1523 KEYSTIDGSLYFEGTKENQAIDMAELITLMPRLQENEOICLYALGETM----- 1574
 QY 115 VYLKDAAHADVNTKEINRQKQEHSGHREGTFRNDGAVVALASGGRYTTDGYIFNAS 174
 DB 1575 -----DNIRSKCPVN-----DSDSTPRT-VPCLCR-----YAMTAE 1606
 QY 175 DIIEDTG-----DAYIVPHGDHYHYI PKNELSASELAEAFLSGRGLNSRTYR 225
 DB 1607 RIARLRSHQVKSVMVCSFPLP---KYHVDGVQKVCCKVLFPDTPV---SVSPRKXA 1660
 QY 226 RQNSDNTSRT-----NWVPSVSNPGTTNTSNTSNTSQAOSQNDISLKKOLYKLP 279
 DB 1661 ASTTHSDRLSGFPLDWM-----TTDSSSTASDTMSLPLOSCTDISIYEPMAPIYV 1712
 QY 280 -SORHVESDGLVDPQAITSRTARGAVAVPHGDHYHFIPYSQWSELEERLARIIPLYRASN 338
 DB 1713 TADVHPEPAGIADLAADVPEPA-----DHV-----DLENPIF---PPRPKXA 1752
 QY 339 HWVPSRPEQSPQPTPEPSPGQPAPNLIK-----DSNSLSVQLVRKVGEGYFEEKG 393
 DB 1753 AVLASRAARVPVAPR-KPTPAFRTAFRNKRLPTFGDPHEHVDALASGITFGDPDVLR 1811
 QY 394 ISR---YVPAKDLPESTVKNLESKQSVSHTLTAKKENAVAPRQOEYDAVYNLLTAA 450
 DB 1812 LGRAGAYITSSDTGS---CHLOOKSVRQ---HNLOCAQLDAVOEKKMPPK---LDTER 1861
 QY 451 HKALF-----XNKGNSDFOALDKLERLNDESTTKKELVDLLAFLAP---ITPR 498
 DB 1862 EKLILLKMQMHPSEANKSRYS-QKVENMKATVVDLTSGARLYTGADVGRIPTAVVRYP 1920
 QY 499 ERLGKNSQIEYTEDEVRIA---QLADKYTTSDBG-IFDEHDIISDEGDAYVTTPMAGS 553
 DB 1921 RGVYSPVTVIERSSPDVALAACNEVLSRNYPTVASQITDEYDAYLDWD----- 1970
 QY 554 HWIGKDSLSDKE---KVAQAAYTKKEGI-----LPSP-----DADVKAN--- 590
 DB 1971 ---GSDSCIDRATFCFAKLRCEYKHAHYHQPTVRSVAVSPFQTLQNLAAATKRCNVY 2027
 QY 591 -----PTGSAALIV-----RVGGER--- 608
 DB 2028 QMRBELPTMS--AVFVVECFKRYACSGEWEYAKOPIRITTENTITTYVTKLGPRAAAL 2085
 QY 609 -----IPVLRLPYVVEHTVEYKNGNLIIPKHQD 636
 DB 2086 FAKTHNLVPLQGVF-MDRFTVDMKRDVKYTPGRTKH 2119

RESULT 21
 US-08-466-277-2
 ; Sequence 2, Application US/08466277
 ; Patent No. 6190666
 ; GENERAL INFORMATION:
 ; APPLICANT: Garoff, Henrik
 ; Lilljestrom, Peter
 ; TITLE OF INVENTION: DNA Expression Systems Based on
 ; Alphaviruses
 ; NUMBER OF SEQUENCES: 27
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Birch, Stewart, Kolach & Birch
 ; STREET: P.O. Box 747
 ; CITY: Falls Church

STATE: Virginia
 COUNTRY: USA
 ZIP: 22040-0747
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/466,277
 FILING DATE: 06-Jun-1995
 CLASSIFICATION: <Unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 07/920,281
 FILING DATE: <Unknown>
 ATTORNEY/AGENT INFORMATION:
 NAME: Murphy Jr., Gerald M.
 REGISTRATION NUMBER: 28,977
 REFERENCE/DOCKET NUMBER: 828-103P
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 703-241-1300
 TELEFAX: 703-241-2848
 TELE: 248345
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2431 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:
 US-08-466-277-2

Query Match 3.4%; Score 141; DB 3; Length 2431;
 Best Local Similarity 19.7%; Pred. No. 0.14;
 Matches 137; Conservative 93; Mismatches 253; Indels 212; Gaps 34;
 QY 56 QGVYVSHGHYHYNG-KVPYDAIISELLMKDPNKLKDEDI VNEVGKGYIKVDGKYY 114
 DB 1523 KEYSTIDGSLYFEGTKENQAIDMAELITLMPRLQENEOICLYALGETM----- 1574
 QY 115 VYLKDAAHADVNTKEINRQKQEHSGHREGTFRNDGAVVALASGGRYTTDGYIFNAS 174
 DB 1575 -----DNIRSKCPVN-----DSDSTPRT-VPCLCR-----YAMTAE 1606
 QY 175 DIIEDTG-----DAYIVPHGDHYHYI PKNELSASELAEAFLSGRGLNSRTYR 225
 DB 1607 RIARLRSHQVKSVMVCSFPLP---KYHVDGVQKVCCKVLFPDTPV---SVSPRKXA 1660
 QY 226 RQNSDNTSRT-----NWVPSVSNPGTTNTSNTSNTSQAOSQNDISLKKOLYKLP 279
 DB 1661 ASTTHSDRLSGFPLDWM-----TTDSSSTASDTMSLPLOSCTDISIYEPMAPIYV 1712
 QY 280 -SORHVESDGLVDPQAITSRTARGAVAVPHGDHYHFIPYSQWSELEERLARIIPLYRASN 338
 DB 1713 TADVHPEPAGIADLAADVPEPA-----DHV-----DLENPIF---PPRPKXA 1752
 QY 339 HWVPSRPEQSPQPTPEPSPGQPAPNLIK-----DSNSLSVQLVRKVGEGYFEEKG 393
 DB 1753 AVLASRAARVPVAPR-KPTPAFRTAFRNKRLPTFGDPHEHVDALASGITFGDPDVLR 1811
 QY 394 ISR---YVPAKDLPESTVKNLESKQSVSHTLTAKKENAVAPRQOEYDAVYNLLTAA 450
 DB 1812 LGRAGAYITSSDTGS---CHLOOKSVRQ---HNLOCAQLDAVOEKKMPPK---LDTER 1861
 QY 451 HKALF-----XNKGNSDFOALDKLERLNDESTTKKELVDLLAFLAP---ITPR 498
 DB 1862 EKLILLKMQMHPSEANKSRYS-QKVENMKATVVDLTSGARLYTGADVGRIPTAVVRYP 1920
 QY 499 ERLGKNSQIEYTEDEVRIA---QLADKYTTSDBG-IFDEHDIISDEGDAYVTTPMAGS 553
 DB 1921 RGVYSPVTVIERSSPDVALAACNEVLSRNYPTVASQITDEYDAYLDWD----- 1970
 QY 554 HWIGKDSLSDKE---KVAQAAYTKKEGI-----LPSP-----DADVKAN--- 590

Db 1971 ---GSDSCIDRATFCPAKLRCPKHAHAYQPTVSASVSPFQNTLQNVLAATKNCNVT 2027
 Qy 591 -----PTGSDAAATN-----RKSEKR--- 608
 Db 2028 QMRELPTMDS--AVENVECFKRYACSGEYMEYAKOPRITTEITTYVTYTKLKGPKAAL 2085
 Qy 609 -----IPLVRLPYVHEHTVEVKNGLIIPHKDH 636
 Db 2086 FAKTHNLVPLQEVF-MDRFTVDMKADVKTPTGKA 2119

RESULT 22
 US-08-714-741-32
 ; Sequence 32, Application US/08714741
 ; Patent No. 6500613
 ; GENERAL INFORMATION:
 ; APPLICANT: Biles, David E.
 ; APPLICANT: McDaniel, Larry S.
 ; APPLICANT: Swiatlo, Edwin
 ; APPLICANT: Yother, Janet
 ; APPLICANT: Crain, Marilyn J.
 ; APPLICANT: Hollingshead, Susan
 ; APPLICANT: Brooks-Walter, Alexis
 ; TITLE OF INVENTION: PNEUMOCOCCAL GENES, PORTIONS THEREOF,
 ; TITLE OF INVENTION: EXPRESSION PRODUCTS THEREFROM, AND USES OF SUCH GENES,
 ; TITLE OF INVENTION: PORTIONS AND PRODUCTS
 ; NUMBER OF SEQUENCES: 47
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Curtis, Morris & Safford, P.C.
 ; STREET: 530 Fifth Avenue
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.
 ; ZIP: 10036
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/714,741
 ; FILING DATE: 16-SEP-1996
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Frommer Esq., William S.
 ; REGISTRATION NUMBER: 25,506
 ; REFERENCE/DOCKET NUMBER: 454312-2460
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 840-3333
 ; TELEFAX: (212) 840-0712
 ; INFORMATION FOR SEQ ID NO: 32:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 8991 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: amino acid
 ; US-08-714-741-32

Query Match 3.4%; Score 141; DB 4; Length 8991;
 Best Local Similarity 18.8%; Pred. No. 1.1; Indels 176; Gaps 19;
 Matches 110; Conservative 70; Mismatches 230; Indels 176; Gaps 19;

Qy 239 PSVSNPGTTNTNTSNNSTNSQASQSNIDS-----LTKQLYKLPLSQRHVESDGL----- 289
 Db 3890 PGGTEPDLDEKASDSNIGALPNQVSDLENQVSLDEVRRLPSDLADTTEGNNVGDYVK 3949
 Qy 290 -----VDPDAQITRTARGAVALPHGDHYHFLPYQOMSELEERIRI 330
 Db 3950 GGLKALATDEKVGILNNTPKALDTAPKALDTALNELGPDGD-----EETPA-- 3995

Qy 331 IPLRYSNHWVDSRPEQSPQPTPEPSPGOPAPANLIKDSNSSLVSQLYKRGEGYVE 390
 Db 3996 -----PAKPEQAPQPKAPAPAKPEKTDQQAED-----YARSEEBY--- 4035
 Qy 391 EKGISRYVFAKDLPESTYKNLESKL---SKOESVS-----HLLTAKENVAPROEFY 440
 Db 4036 -----NRLPQOQPPKAKEKPAPAPKEQVPYAPGWSMRTLLAPRDLARQAELIA 4085
 Qy 441 DKAYNLITFAKAL--FANKRNSD-----FOALDKLRLINDESTNKEKLYDILLAF 491
 Db 4086 QKQ-----TELGLKLDSDLPBEGKTQDELDEKAGEABLDKKAAGLPRKASDLEKEISNLEIL 4141
 Qy 492 LA-----PITHEBRLGKPSQIETYEDEVRIAOQLADKYTTSDGYIFDEHDIISDEGDA 544
 Db 4142 LGADSEBDTALNKLATKKAELKTKQKELDAL-----NELGPDGE 4185
 Qy 545 YVTPHMGSHWIGKDSLSDEKVAOAATYKKGILPSPDADVQANPTGSDAAATYKRVK 604
 Db 4186 EETP-----AAPQPEQAPAPAPKEQPTPAPKEQPTPAPKEQAPAPA---PK 4230
 Qy 605 GEKRIPLVR-----LPYVHEHTVEVKNGLIIPHKDHYNIKFAMPDDHTYKAPNGYTL 658
 Db 4231 PEQAPAPKEQAPAPAPKEQPTGPKI EELLLEK-----AGL 4269
 Qy 659 EDLPATTKYVEHPDERPHSNDGNASEHYLGKDHSEDPNKNPKADEBEVEETPAPE 718
 Db 4270 GKAGADLKEAVNNEGGS-----AGPSQPEEBAEAPAPKEQTE--PTQPE 4313
 Qy 719 VPQVETKEVKAQLKEAEVLLAKVTDSUKANATETTLAQLNNLTQ 764
 Db 4314 EPAGETAPKPEKPAKOPKAKEKTDQQAEDYARRSEEBYRLTQO 4359

RESULT 23
 US-08-976-255-14
 ; Sequence 14, Application US/08976255
 ; Patent No. 6136581
 ; GENERAL INFORMATION:
 ; APPLICANT: Jono, Keith E.
 ; APPLICANT: Plozman, Gregory
 ; TITLE OF INVENTION: KINASE GENES AND USES
 ; NUMBER OF SEQUENCES: 53
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Lyon & Lyon
 ; STREET: 633 West Fifth Street
 ; STREET: Suite 4700
 ; CITY: Los Angeles
 ; STATE: California
 ; COUNTRY: U.S.A.
 ; ZIP: 90071-2066
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
 ; MEDIUM TYPE: storage
 ; COMPUTER: IBM compatible
 ; OPERATING SYSTEM: IBM P.C. DOS 5.0
 ; SOFTWARE: Fast-Seq for Windows 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/976,255
 ; FILING DATE: No. 6136581ember 21, 1997
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 60/031,675
 ; FILING DATE: No. 6136581ember 22, 1996
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Wardburg, Richard J.
 ; REGISTRATION NUMBER: 32,327
 ; REFERENCE/DOCKET NUMBER: 229/182
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (213) 488-1600
 ; TELEFAX: (213) 955-0440
 ; TELE: 67-3510
 ; INFORMATION FOR SEQ ID NO: 14:
 ; SEQUENCE CHARACTERISTICS:

LENGTH: 1503 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: Protein
 US-08-976-255-14

Query Match 3.3%; Score 138; DB 3; Length 1503;
 Best Local Similarity 18.3%; Pred. No. 0.11;
 Matches 168; Conservative 127; Mismatches 293; Indels 329; Gaps 43;

```

QY 14 ENNRVYINGKATQCTENTLTPDEVSKREGINA-----EOLYIKITD 55
DB 413 OSQRDEVDVE---QQMNAIKPNTNSRDSSNNAFPLDHPADRILGREMEEVLTETRS 469
QY 56 OG---YVTSHGDPHYH---YNGKPYDAIIE---ELMKDPYKUKD---EDI 97
DB 470 QGLSFYVWEAAGHDFDEKSRGHLDEGLSYTIFYPVVEFESSLSDPGPKODSGQDV 529
QY 98 VNEVKG---GYVIKVDGKYVYVVKDAHA-----DN-VRTKEEINQ 135
DB 530 PLRVPVYVYFDHNLNLSVGSDDYIQLEKSGSLIEDYPPALLTTMDNPERKGPESQL 589
QY 136 KOEHSQHREGGTPRNDGAVALARSQGRYTTDDGYIFNASDIEDT--GDAYIV---PH 188
DB 590 TALRS-----VELEES---STDEDFQSSIDPDKOSLPGDLHVTSGPESPF 632
QY 189 GGHYHPIKNE-----199
DB 633 NMFNDVDSDELPSHOKIFDLMELNGVADFKPATLSSLDNPKESVITGHEKEKPRK 692
QY 200 -----LSASELAALAEAFLSGRGNLSNRTYRRQNSDNTSRTNVY 238
DB 693 IFDSEPLCLDNLMHODNFDPLNVGELSENFLFLOEKNLKGLSKSEKHINDQT----- 747
QY 239 PSVSNPGTT---NTNTSNNSNTNSQASQNDIDSLKOLYKLPISQHVESDG----- 288
DB 748 -ELKNAAGTEAMLETSCRNSLDTELQFAENKPGLSIL-----QENVSTKGDQTDVW 797
QY 289 LVFDPKQITSRTAGVAVPHGDHYHFIYSGMSELEERARIITPLRYRNNHWPDPRQ 348
DB 798 LGDGLTSLQSSPEVQVP-----PTS--PETETPRPV-----PPDGLPTQ 837
QY 349 PEPQPT-----PEP-----SPQPAAPNLKIDNSSLSVQLVRKVGEGYVEEKGISR 396
DB 838 GETQPCLDVIYEPEDLHODISPDATVP--VELISTDARTHSIDNR-----882
QY 397 YVFAKDLPSFTVKNLESKLSKQESV--SHLTAKKENVAPR---DQFYDKAYNLTTEAH 451
DB 883 ---SQDSPESBEETL--RLTESDVLADILASRVSGSLPELGQELHNKPFSEEDHSH 937
QY 452 KALFXKKGKNSDFQALDK-----LLEKLNDESNKEKLVDDLLAFAPITHPERLKG 503
DB 938 RLEKLEAVEETLNQNSDKAAKEAGLVSAISDSQTSODLSLEDSASAPFPASEPS--LET 996
QY 504 PMSQLEYTDEYRIAQLADKYTTSQGYIFDEHDIIDEGDAYVTPMHGSHWIGKXSLSD 563
DB 997 PDS-LESV-----DVEALDLSLGSHTP-----OKLVP 1023
QY 564 KERVAQAAYTKE-----KGILPPSPDADYKANP--TGDSAAALYNKVEKRIPLVRLPY 616
DB 1024 PKRPADSGYETENLJESPEWTLHPAREGTADSEBATTGGDGHs-----GLPRPVY---- 1073
QY 617 MEVHTVEVXKGNLIITHKHYNHIFKAMPEDHTYKAPNGYITLEDLPAITIKYVYEHDEPR 676
DB 1074 VTSADAGHGRGTEVT-----ETFTAGSQGSYRD-----SAFSDSDSEPR 1113
QY 677 HSNDSGNASEHVLGKSDSEDNKNFKADEBEVEETPAPEPVQYETEKVEAKLEAF-- 735
DB 1114 EKR-----SEEVPG-----TSPSALVIVQEQPLPEPVLPBGSPAQDSCLEARKSQPDE 1162
QY 736 -VLLAKVTDSLKANAT 751

```

DB 1163 SCLSAHNSDLELRAT 1179

RESULT 24

US-09-513-783A-22
 ; Sequence 22, Application US/09513783A

Patent No. 6416959

GENERAL INFORMATION:
 ; APPLICANT: Giuliano, Kenneth A.

APPLICANT: Kapur, Ravi

TITLE OF INVENTION: A System for Cell Based Screening

FILE REFERENCE: 97-022-LI

CURRENT APPLICATION NUMBER: US/09/513,783A

CURRENT FILING DATE: 2000-02-25

NUMBER OF SEQ ID NOS: 180

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 22

LENGTH: 1610

TYPE: PR

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence:

OTHER INFORMATION: EYFP-DEV4-MAP4-BFP construct

US-09-513-783A-22

Query Match 3.3%; Score 137.5; DB 4; Length 1610;
 Best Local Similarity 19.8%; Pred. No. 0.14;
 Matches 174; Conservative 103; Mismatches 337; Indels 265; Gaps 41;

```

QY 53 ITDQGY-----VTSHGDPHYHYNGKPYDAIIESEELMKDP--NYKLDE-----DIV 98
DB 62 VTFPGYGLQCFARYPDMMKQNDFFKSAWBEQYVQRTTFFKQDGVYKTRAEVKFEGDTLV 121
QY 99 NEVK--GGYVYKDG-----KYVYVVKDAHADNVTKKEINRQKEHS-----Q 141
DB 122 NRIELKGIIDFKEDGILGKLEYNVNSHNVYIMADOKKNGIKVNFKIRINIEDGSVQLAD 181
QY 142 HEEGTPRNDGAV-----ALARSQ-----RYTDDGY-----IFNAS 174
DB 182 HYQNTPTIGDGVLLPDNHYLSYQSALSQDPNEKRDHWLLEFVTAAGITGLMDELYKKG 241
QY 175 DIIEPTGAYIVPHGDHYHYP--KNELASASELAALAE-----PLS 213
DB 242 DEVQGAALSLV---DALTEPPEIEGEEKRDFMALEAPYDDIVGETVEKTEFIPILD 298
QY 214 GGNLSNSRTYRRQNSDNTSRTNVPSVSNPGTNTNTNSNTNSQASQNDIDSLKQ 273
DB 299 GDEKTNSESKKKPKCID--TSQVEGIPs--SKP--TLANGDHGMEGNNTAGSPPTD---FLEE 352
QY 274 LVKLP---LSORHVESDGLVFPQAQITSRTAGVAVPHGDHYHFIYSGMSELEERARI 330
DB 353 RVDYDPYQSSQWPEDEASCFQPOVLDTQ---AEFNEH-----RDGLADL 398
QY 331 IPLRYRNNHWPDPRQSPQTPPEPQPAAPNLKIDNSSLSVQLVRKVGEGYVE 390
DB 399 LTVSSGPTNASSAFTERDNP-----EDSYGMPLPCOSF---ASTAVSQGSWSGAPNSPCS 450
QY 391 EKGISRYVFAKDLPSFTVKNLE-----SKLSQESVSHTLTAKKENVAPRDOEFYDKAYN 445
DB 451 BECVS-----PEVTIETLOPATELSKAAVESVKEOLPAK-----485
QY 446 LITTEHKAFLXKKGNSDFQALDKLERLND-----ESTNKEKLVDDLLAFAPITHPERL 501
DB 486 -----ALETNAEBQTTDVVHSPSTDTTGTGPTBEALADIE-----520
QY 502 GKNPQSOLEYTEDEVRIAQLADKYTTSQGYIFDEHDIISDEGDAY-----545
DB 521 -----ETTKPDVILANVTQSTESDMFLAQDMELTQTEAAHANNIITLPTPEPSSYK 573
QY 546 -VTPMHGSHWIGKXSLSDKEKYAA-----QAYTEKGIPLPPSPDADYKANPT 592
DB 574 DVAPMEEBEIVGNDJTSKETETETTLPIKMDLAPEDVLLTKETELAP---AKGVSLIS 629

```

QY 593 GDSAAAIYNRVKGKRIPLVRLPYVNEHTVEVKGNNLIIPKHQHYHNKIFAMFDDHTYKA 652
DB 630 EIEBALAKNDVRSAE-IPVAQ-ETVASEFEVLVATEVLP----- 667
QY 653 PNGYLEDLFAITIKYVNEHP--DERPHSNDGWN-ASEHVLGKKD-HSEDPKNKFADEE 708
DB 668 -----SPITTLITLDVTLPLEAEKPLVDTMPSLETETMLGKETAPPTETLGNKXKMS 721
QY 709 PVEETPAE-----PEVPQVETEKVEAQLKEAEVLLAKYTTSSILKANATELGLANN 760
DB 722 PLPESEVTLGKDVVLPLETKVAEFNNV--TPLSEEEV--TSVKDMSPSA---ETEAPLAKN 775
QY 761 LTLQ-----IMDNNSIMAEKLLALLKGSNPSVSKEK 794
DB 776 ADLHSGTELIIVDNS--MAPASDLALPLETKVATVPIKOK 812

RESULT 25

US-08-447-031A-2
Sequence 2, Application US/08447031A

Patent No. 5851794

GENERAL INFORMATION:

APPLICANT: GUS, Bengt

APPLICANT: HOOK, Magnus

APPLICANT: JONSSON, Hans

APPLICANT: LINDBERG, Martin

APPLICANT: PATTI, Joseph

APPLICANT: SIGMANS, Christer

APPLICANT: SWITALSKI, Lech

TITLE OF INVENTION: A COLLAGEN BINDING PROTEIN AS WELL AS

NUMBER OF SEQUENCES: 8

TITLE OF INVENTION: ITS PREPARATION

CORRESPONDENCE ADDRESS:

ADDRESSEE: Burns, Doane, Swecker & Mathis

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/447,031A

FILING DATE: 22-MAY-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/861,804

FILING DATE: 21-AUG-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO PCT/SE91/00707

FILING DATE: 22-OCT-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: SE 9003374-7

FILING DATE: 22-OCT-1990

ATTORNEY/AGENT INFORMATION:

NAME: McGowan, Malcolm K.

REGISTRATION NUMBER: 39,300

REFERENCE/DOCKET NUMBER: 012889-006

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1183 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-447-031A-2

Query Match 3.3%; Score 136.5; DB 2; Length 1183;

Best Local Similarity 19.2%; Pred. No. 0.1;
Matches 189; Conservative 120; Mismatches 317; Indels 357; Gaps 54;

QY 15 NNRYSYIDGKQATQKNTNLPDEVSKEKGINAEQIVKITDQGVTHSGHGHYHYNGKVP 74
DB 195 NNRKSYV-SKDIITK-----DOIQGOQDLSTLNTINVT-----GHSNYSQO-- 237
QY 75 YDAIIEELMKDKDNYKLEKEDIVNEVK-----GGY-----VIKVDKYYVYL 117
DB 238 -SALTDEPKAF--FGSKITVNTKNTIDVITPQGYGYSNSISYKTKITNEQKKEFPNN 294
QY 118 KDAAHADNVTKEEINRQKQESQHREGGTPRNDGAV-----ALASQGR 162
DB 295 SQAWYQEH--GKEEVNGKSFHHVTHVINANAGISGTVKGELKVLKODKDTKAPIANVFK 352
QY 163 YTTDDGYIFNMS-----DIIED-TGDAYI--VPHGDHYYIPKNSLSASELAAPLSGR 215
DB 353 LSKKDGSVVDKQKEIITDANGIANIKALPSGD--YIUK-----EIEAPRPYFPDK 403
QY 216 GNLSNSRTYRQNSDNTSRTWVPSVSNPGTTNTNTSNNSTNSQASQSDIDSLKQLY 275
DB 404 ---DKEYPFTMKDIDNGYF-----TTINAKAIEKTKVSAOK----- 439
QY 276 KLPLSQRHVESDGLVFPDPAQITSTARGAVAPHGDHYHFIPIY--SQNSELEERARILPL 333
DB 440 -----VMEGTQKVKPTIYFKLYKQDNDONTTPVDKAEIKLEDGTTKYV-- 483
QY 334 RYRNNHWVPSRPPROPSPQRPPEPSPGQAPRPNKIKISNSLSVQLVRKV-----E 385
DB 484 -----W--SNLPN-----DRNGKAIYLVKEVVAQGEDPTPE 514
QY 386 GYVFEKGIISRYVPAKDLPSETVKNLESKL-----SKQESVSHLTAKKENVAPRD 436
DB 515 GYTKKENGL--VANTTEKPIETTSISEGKAWDDKDNQDKRPEKVSNNLANGSKVTLTD 572
QY 437 -----QEF-----YDKAVNL--LTEAH-----KALFXNK----- 458
DB 573 VTSETNWKYEFKDLPRKYDEGKKIETVTEHDVHKYTTDINDGTTITNKYTPGETSATVTKN 632
QY 459 ---GRNSDFQALDULRL--NDESTNKEKLVDDLLFLAITIPERLGRKNSQIEVTEED 513
DB 633 WDDNNNDGKRPTIKVELYODGKATGKTALINESNNWTHWTGULDEKAK--GQGVKRYTVE 691
QY 514 EVRIAQLADKYTTSDDYI--FDEHD-----IISDEGDAVYTPHMGSHWIGKDS----- 560
DB 692 EL-----TKXKGYTHVDNNDMGNLIVTNKYTPETTSISEGKAWDDKDNQDKRPE 742
QY 561 -----LSDKEKVAQAATYKEKGLPPSPDADVYKAPPTGDSAAAIYNRVKGKRIPLVRL 614
DB 743 KVSNNLADGKVKTLDVTSF-----TNWKYEFKDLPRYD-----EGKKIEYTVTE 788
QY 615 PYMVEHYVEVNGNLI-----IPKHQHYN-----IKFAFPD----- 647
DB 789 DHVKDYTTDI-NGTTITNKYTPGETSATVTKNWDNNNDGKRPELIKVELLYQOKATGK 847
QY 648 -----HTY-----KAPN--GYTLEDLFAITIKYVNEHDERPHSNDGWN----- 684
DB 848 TAILNESNNWTHYTGULDEKAKGQGVKYVDEL--TKXKGYTHYD-----NNDGNLIVT 901
QY 685 -----ASEHVLGKKHSED--PNK--NFKADEPVEETPADE-----VP 720
DB 902 NKYTPEITTSISGEKAWDDKDNQDKRPEKVSNNLANGKVKTLDVITSETMWKYEFKDLPL 961
QY 721 QV-ETEKVEAQLKEAEVLLAKYTTSSILKANAT-----ETLAGLRNNLTLLIMDNN 770
DB 962 KYDEGKKIEYTVTEDEH-----VKDYTTDINGTTITNKYTPGETSATVTKNWD--DNNN 1012
QY 771 IMAEAKLLALLKGSNPSVSKEK 793
DB 1013 -----QDKRPTEIKVE 1024

RESULT 26


```

Db 490 -----ALETMAEQTTDVHSPSTDTTGGPDTEALAKDIE- 524
Qy 498 PERLGKPNQIETEDEVRIAOADKYTTSDDGYIFDEHDIISDEGDY----- 545
Db 525 -----EITPEVILANVTOPSTESDMFLADQMDLITGTAAHANNIILTEPDE 573
Qy 546 -----VTPHMGSHWIGKDSLSDKEKVA-----QAYTEKGIPLPSPDADVK 588
Db 574 SSTKVAPMEEEIYVGNDDTSPKTEITETTLPIKMDLAPPEVDLIRKETELAP-----AKGM 629
Qy 589 ANPTGDSAAAIYNRKYGKEKRIPLVLPVVEHTVEYKGNLIIPKDHNNIKFAMFDDH 648
Db 630 VLSLEIEBALAKNDVRSAE-IPVAQ-ETVASETEVVLATEVLP----- 671
Qy 649 TYKANGYLTLEDLPATIKYVEHP--DERPHSNDGKN-ASEHVLGKGD-HSEDPNKAFK 704
Db 672 -----SDPITTLTKDVTLPLEAERPLVTDMPLEIEMTIGKETAPTEENLGM 721
Qy 705 ADEEVEETPAE-----PEVPOVETEKVEAOKEAEVLLAKYTDLSLKANATELLAG 756
Db 722 KDMSPLPSEBVTLGKDVILPETKVAEFNNV-TPLSEBEV--TSVKDMSPSA---ETEAR 775
Qy 757 LRNNLTQ-----INDNSIMAEKELALLKGNPSVSKEX 794
Db 776 LAKNADLHSGTELIYDNS--MAPASDLALPLETKVATVPIDKX 816

RESULT 28
US-08-015-973-1
; Sequence 1, Application US/08015973
; Patent No. 5604094
; GENERAL INFORMATION:
; APPLICANT: Schlessinger, Joseph
; TITLE OF INVENTION: A NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESS: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/015,973
; FILING DATE: 10-FEB-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mastrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 7683-021
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)790-9090
; TELEFAX: (212)869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2308 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-015-973-1

Query Match 3.2%; Score 134; DB 1; Length 2308;
Best Local Similarity 19.2%; Pred. No. 0.46;
Matches 169; Conservative 123; Mismatches 276; Indels 310; Gaps 44;
Qy 51 IKITDGYVTSHGHDVHYNGKVPYDAIISSELLMKDPYKLDKEDIIVNEVKGVIKVD 110

```

```

Db 954 VGVTYQGSILFGSPSHI-----PIPKSLITPTASLQPTHALSG-----D 993
Qy 111 GKYYVYLKDAHANVNRKEIRINQKQESHQRRGGTFRNDGAVALL-----ARSGRYT 164
Db 994 GEW-----SGASSDS-----BFLPDTDGLTALNISSVSAEFTYT 1030
Qy 165 T--DDGYIPNADDIETDGAIVPHGDHYH-----IPK-----NELSASELAAE 209
Db 1031 TSVFGDDKALKSKSEIITGNTELIQIPFNEMVYPSSEVTWPNWYDNNKLNAGIQETSV 1090
Qy 210 AFLSGRG-----NLSNRYR-RQNSDNTS-----RTNWVP 239
Db 1091 SISSTKGMFPSSLATTTKVYFDHEISQVPENNFVOPHTVSOAGDSTSLPVLANSNP 1150
Qy 240 SVSNPQTT-----NTNNSNNTSQAAS-OSNDISLLKOLYKPLSGRHYES 286
Db 1151 ASSDPASSEMLSPSTOLLIFETYSASFTEVLLQPSFOASDVDTLKTIV--LPA---VPS 1204
Qy 287 DGLVFPDPAQ-----ITSRTARGVAVPHGDHYFIPIYSOMSELEERIANIIPURY 335
Db 1205 DPLIVETPKVDKISITMHLIVSNASASENNLHSTSVFPDVSPTSHHSHASLGGLTISY 1264
Qy 336 RSNHWVPDSRPEQSPQPTPEPSPGQAPNLKIDSNSLSVQLVRKYGEGYVEBKGIS 395
Db 1265 ASEKYEPVILTKSESSHQVVP-----SLYSNDELFOQTANLEINQAH--PPKG-- 1308
Qy 396 RYVFAKDLPS--EIVKULESKTSKQESVSHTLTAKKEVAVR-----DOEFYDAX 444
Db 1309 RHVFATPVLSTIDELNLTINKLIHSDER--LTSKSSVTGKVPAGIPTVASDTPFVSTDH 1365
Qy 445 NL-LTEAH-----KALFNKKG----- 459
Db 1366 SVPIGNHGVAITAVSPRRDGSVTSTKLLFPKATSELSHAKSDAGLVGGEDEGDTDDG 1425
Qy 460 -----RNSDFOALDKLIERLNDESTNKEKLVDDLAFIATYPERLGKNSQIET--- 510
Db 1426 DDDDDRSDDLSTIHKCMS--CSSYRESQEKVMD-----SDTHENSIMDONNISYSLS 1478
Qy 511 --TEDEVRIAOADKYTT-----SDGYIFDEHDI-----ISDE 541
Db 1479 ENSEEDNKNVTSVSSDSQGMDRSPGKSPSANGLSQKHNDGK--BENDIQOTGASALLPLSPE 1536
Qy 542 GDAYVTPHMGSHWIGK---DLSLSDKEKVA--AQAYTEK--GIL-----PPS 582
Db 1537 SKANAVALTSDSESGCGTSTSLNENETSTDFADTNEKXADGILAAAGDEITPGFQOS 1596
Qy 583 PDADV-----KANPTGDSAAAIYNRKYGKEK--IPVRLPYVVEHTVEYKN 626
Db 1597 PTSSVTSENSEVFPVSEAEASNSHESRIGLAEGLESEKKAVIDPLIVYSALTFLICLVLY 1656
Qy 627 GNLIIPHK-----DHYNNIKFAMFPDDHY-----KAPNGTLEDLFA-TIKYVEHPDE 674
Db 1657 GILYWRKCFQTAHF-----YLEDDTSPRVISTPPTPIPIPSIDVGAIPIKHPPKHVAD 1710
Qy 675 RPHSNDGKNASEHV-----LG-----KQDSEDPNKN 702
Db 1711 L-HASSGTFEEFEVQGVDTVDLGITADSSNHPDKKHKN 1747

RESULT 29
US-08-448-164-1
; Sequence 1, Application US/08448164
; Patent No. 5925536
; GENERAL INFORMATION:
; APPLICANT: Schlessinger, Joseph
; TITLE OF INVENTION: A NOVEL RECEPTOR-TYPE PHOSPHOTYROSINE
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESS: PENNIE & EDMONDS
; STREET: 1155 Avenue of the Americas
; CITY: New York

```

```

STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/448,164
FILING DATE: 24-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/015,973
FILING DATE: 10-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Mierock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7683-021
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2308 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-448-164-1

```

Query Match 3.2%; Score 134; DB 2; Length 2308;

Best Local Similarity 19.2%; Pred. No. 0.46; Matches 169; Conservative 123; Mismatches 276; Indels 310; Gaps 44;

```

51 IKITDGGVYTSKGDHYYNGKVPYDAIISEELMKDPNYKLDKEDIVNEVGKGYIKVD 110
954 VGVTYOGSLFSGPSHI-----PIPKSLIPPTASLQPTHALSG-----D 993
111 GKYYVYLKAAHADVNRKTEINRQKQESHREGGPRNDGAVL-----ARSGRYT 164
994 GEW-----SGASSDS-----EFLPDTDGLTALNISPVSAEFTYT 1030
165 T-----DDGYIFNSDIETDGDYIVPHGDHYH-----IPK-----NELSASELA 209
1031 TSVFGDDNKALSSSEIIYGNETELQIPSEMYTPSESTWPMYNDVANKLNSLOETSV 1090
210 AFLSGRG-----NLNSRTYR-RQNSDNTS-----RTNWVP 239
1091 SISSTGMPFPGSLAHTTKVFDEHISQVFNPNFSVQPTHVVSQASGDTSLKPVLSANSEP 1150
240 SVNRPQT-----NTNTSNNSNTNSQAS-OSNDIDSLKOLYKLPISQKHVES 286
1151 ASSDPASSSEMLSPSTQLLFYETSAFSTEVLLQPSFQASVDVTLTKTV-----LPA 1204
287 DGLVFPDPAQ-----ITSRTARGVAVPHGDHYHPIPYGOMSELEERIRARIIPRAY 335
1205 DPLIVETPKVDKISSITMLHIVSNSASSEMLHISTVPVDVSTFTHMSHSAISLQGLTISV 1264
336 RSNHWVPDSRREPQSPQTPREPSPGPAPAPNLKIDSNSSLVSQLVRKVGEGVFEKGIS 395
1265 ASKEXEYVLLKSSSHQVVP-----SLYSDELFOPTANLEINQAH--PPRG-- 1308
396 RYVFAKDLPS--ETVKLLESKLSKQESVSHTLTAKKENVAPR-----DOEFYDKAY 444
1309 RHVFAPVPLVIDEPLNTLTKLHSDI--LTSKSSVYGVKVFAGIPVVASPTFVSTDH 1365
445 NL-LTEAH-----KALFXNKG----- 459
1366 SVFIGNCHVAITVSPRDSVSTKLLFSPKATSEISHAKSDAGLVGGEGDGTDDG 1425
460 -----RNSDFQALDKLLERLNDSTNKEKLVDDLAFIAPITHPERLGRNSQIEY--- 510

```

```

DB 1426 DDDDDRGSDGLSIHKNS-CSSYRESQEKVNM-----SDTHENSLMDONNPISYSLS 1478
QY 511 --TEDEVRIAOLADKYT-----SDGYIFDEHDI-----ISDE 541
DB 1479 ENSEEDNRVTSVSSQGTMDRSFGKSPSANGLSQKHNGK--EENDIOTGSALPLSDE 1536
QY 542 GDAYTTPHMGSHWIGK--DSLSDKEKVA--AOAYTKER---GIL-----PPS 582
DB 1537 SKAMAVLTDEBSGSGQGTSDSLNENETSTDPSPADTNEKADGILAAAGDSEITPGFQS 1596
QY 583 PDAV-----KANPTGDSAAATYRNVKGER--IPVRLPYMVEHVEYEVKN 626
DB 1597 PTVSVTSENSEYFHVSEARASNSHESRIGLAEGESEKKAIVPIVVSALFTICLVLY 1656
QY 627 GNLIIPHK-----DHYNIKFAWPDHPT-----KAPGYTLEDLFA-TIKYVEHPDE 674
DB 1657 GILTYWRKCFOTAHF-----YLEDSTSRVYSTPTPIFPPISDVGALPIKHPFHVAD 1710
QY 675 RPHSNDGWMGNASEHV-----LG-----KKHSEDPKN 702
DB 1711 L-HASSGFTEEFBEVQSVTDVIGITADSGNHDPNKHKN 1747

```

RESULT 30

US-08-081-929-2
Sequence 2, Application US/08081929
Patent No. 6160090

GENERAL INFORMATION:

APPLICANT: Schlesinger, Joseph
APPLICANT: Barnea, Gildad
APPLICANT: Grumet, Martin H.
APPLICANT: Margolis, Richard U.
TITLE OF INVENTION: A NEW CLASS OF RPTASES: THEIR
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/081,929
FILING DATE: 23-JUN-1993
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Coruzzi, Laura A.
REGISTRATION NUMBER: 30742
REFERENCE/DOCKET NUMBER: 7683-041-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212 790-9090
TELEFAX: 212 869-8864/9741
TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2308 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-081-929-2

Query Match 3.2%; Score 134; DB 3; Length 2308;

Best Local Similarity 19.2%; Pred. No. 0.46; Matches 169; Conservative 123; Mismatches 276; Indels 310; Gaps 44;

```

51 IKITDGGVYTSKGDHYYNGKVPYDAIISEELMKDPNYKLDKEDIVNEVGKGYIKVD 110

```

```

Db      954 VGVTVQSLFSGPSHI-----PIKSSLIPTASLIQPHALSG-----D 993
Qy      111 GKYIYVLDAAADAVNRTEELINROKQEHSHREGTFRNDQAVL-----ARSGRYT 164
Db      994 GSW-----SGASSDS-----EFLLPDTDLTALNISPVSVAEFTYT 1030
Qy      165 T-----DDGYIFNASDIIEDTDAYIVPHGDHYH-----IPK-----NELSASELAAM 209
Db      1031 TSVFGDDNKALSKSEITIGNETELQIPSPENMAYBSESTVMNMNDNNKLNASIQEHSV 1090
Qy      210 AFLSGRG-----NLSNRTYR-RQNSDNTS-----RTNWP 239
Db      1091 SISSTKGMPCSLAHTTKVPDHEISQVPENNFVSQPTHTVQAAGDTSILKPVLSANSP 1150
Qy      240 SVSNNGTT-----NTNNSNTNSQAS-QSNDISLKLQYLPLSQRHES 286
Db      1151 ASSDPASSSEMLSPSTQLLFYETSASFTEVLLQPSFOASDVDTLKTVA--LPA--VPS 1204
Qy      287 DGLVDPDAQ-----ITSRTARGVAVPHGDHYHPIPSQMSLEERLARIILPURY 335
Db      1205 DPLVETPKVDKISSTMLHLIVSNBASSENMLHSTSPVFDVSPSHMHSASLQGLTISY 1264
Qy      336 RSNHWVPDSRPQSPQPTPEBSPGPQAPANIKIDSNSLVSQLVKRYGEGVFEKGIS 395
Db      1265 ASEKEPEVLLKSESSHQVVP-----SLVSNDELFOGTANLEINQAH--PPKG-- 1308
Qy      396 RYVPAKDLPS--EYKAKLESKLSKQESVSHITAKKENAVP-----DQEFYDAY 444
Db      1309 RHVFATPVLISIDEPINTLKLHSDI--LSTKTSVTKGVAGIPTVADTEVSDH 1365
Qy      445 NL-LTEAH-----KALFXNG----- 459
Db      1366 SVPIGNGVAITAVSPHRDGSVSTKLLFPFSKATSELSAKSDAGLVGGGSDPTDDG 1425
Qy      460 -----RNSDFQALDKLERLNDESTNKEKLVDDLAFAPITPBERLQKPSQIEY-- 510
Db      1426 DDDDDRGSDGLSIHKMS-CSSYRESQEKVMD-----SDTHNSLMDQNNPSTSYLS 1478
Qy      511 --TEDEVRIQLADKYTT-----SDGYTFDEHD-----ISDE 541
Db      1479 ENSEEDNRVTVSOSQTMDRSPKSPSANGLSQKHNDGK--BENDIQTSGLALLPLSPE 1536
Qy      542 GDAYVTPMHGSHWIGK--DLSLQKEKA--AQAYTKEK--GIL-----PPS 582
Db      1537 SKAMAVLTSDESGGQGTSDSLNENETSDSPADTNEKADGILAAQDSITTPFPOS 1596
Qy      583 PDADV-----KANPTGDSAAAIYNRVYGEKR--IPLVRLPYVVEHTVEYKN 626
Db      1597 PTSSVTSENSEVFHVSEAAASNSHESRIGLAEGESEKKAIVIPVIVSALFICLVLV 1656
Qy      627 GNLIIPHK---DHYNIKFAMFDHTY-----KAPNGYTLDELFA-TIYYVEHDE 674
Db      1657 GILYWRCKFCOTAFH-----YLBSTSPRIVISTPPTPIFPISDVGAIPIGHPKHAD 1710
Qy      675 RPHSNDGNGNASEHY-----LG-----KKDSEDEPNKN 702
Db      1711 L-HASSGTEFEFEVQCTVDLGITADSNHNDKNKN 1747

RESULT 31
US-09-599-652-2
; Sequence 2, Application US/09599652
; Patent No. RES7741
; GENERAL INFORMATION:
; APPLICANT: HOSTETTER, MARGARET K.
; APPLICANT: GALE, CHERYL A.
; APPLICANT: BENDEL, CATHERINE M.
; APPLICANT: TRO, NIAN-TUN
; APPLICANT: KENDRICK, KATHLEEN
; TITLE OF INVENTION: CANDIDA ALBICANS GENE, INTEGRIN-LIKE
; NUMBER OF SEQUENCES: 12

```

```

CORRESPONDENCE ADDRESS:
ADDRESS: MUEITING, RAASCH, GEBHARDT & SCHWAPPACH, P. A.
STREET: 119 NORTH FOURTH STREET, SUITE 203
CITY: MINNEAPOLIS
STATE: MINNESOTA
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/599,652
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/642,846
FILING DATE: 03-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: MUEITING, ANN M.
REGISTRATION NUMBER: 33,977
REFERENCE/DOCKET NUMBER: 110.00280101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1217
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1664 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-599-652-2

Query Match      3.2%; Score 133.5; DB 1; Length 1664;
Best Local Similarity 17.9%; Pred. No. 0.3;
Matches 159; Conservative 128; Mismatches 364; Indels 237; Gaps 37;

Qy      12 VKENNRVSYIDG-----KATQKTEMLTDEYKESKEGIAEQIVKITQGVYTSIGD 64
Db      140 VNEHAPARYINTSPNKSIMKKATPKA--SPKVA-----FTVNPE 178
Qy      65 HYHYVNGVVPVD-----AIISEELMDPN-YKLKDEIVNEVGKGVYIKVDG 111
Db      179 IHHIPDRNVEEEDSOQKEDSVPEPLIOH--WKDPSQFNYSDEDTNASTVPTPIHTTK 236
Qy      112 KYVYVLDAAADAVNRTEELINROKQ--EHSQHREGTFRNDQAVLARSQRY--TTD 166
Db      237 PTFQQL-----LNKNNEVNSEPEALTDKLRKRE-----NFSNLSIDEKYNLYLSPTN 283
Qy      167 DGYIFNASDI--IEDTGDAYIVPHGDHYH-----PKNELS--ASELAAAEFLSGR 215
Db      284 NNNKKNVSDMDSHLQONLQDASKNTNENIHLSPALKAQKDIENPLNSLTNADISLSS 343
Qy      216 GNLSNRTYRQNSDNTSRTNWPVSVPNPTNTNTSNNSTNSQASQSDNDISLKKQ-- 273
Db      344 GSSQSSLSLQSLANDNRVLESVSGSKVNPGL-----SLNDGKIGPSDEYVESLPPDL 396
Qy      274 -LYKLPLSQRI--VESDGLVDPDAQITSRTARGVAVPHGDHYHPIPSQMSLEERLARI 330
Db      397 SRDLKLETTKEHDAEHNENETIDAKSTNTNKGQLVSDHLDSEFDRS--YHHTQSLINTL 455
Qy      331 IPLAIRYS---NHWVPDSRPQSPQPTPEBSPGPQAPANIKI---DSNSLVSQLVK 382
Db      456 LNSAQOSQISLNALEKOROTEOQOTQAAPEEBEETSPDNITKVQOEPKSNLEFVKVIK 515
Qy      383 VGEYVFEKGISYVFAKDLPSETVKNLESKLSKQESVSHITAKKENAVAPRDOEFDK 442
Db      516 -----EPVSATETKAKKEFSSILIKWEDETLAEAPADHFK 552
Qy      443 AVNLTTEAHKALFYXNGRNSDFQALDKLER--LNDESTNKEKLVDDLAFAPITPBERL 501

```

```

Db      553 KEN---EA-----NSHVEDTDLALKKALNDEESDTJONSTKMSIRPHISDWMK 599
Qy      502 GKPNQIEYTEDEVRIAQLADKYTTSQGYIFDEHDIISDEGAYVTPHMGHSHWIGDSL 561
Db      600 EDSND-----GDRDNDDISFEKSDILNDVSGT-----SDII 632
Qy      562 SDKEKVAQAQYKKEGILPPSPDADYKANFTGDSAAIYNRKGEKRIPLVRLPYWEHT 621
Db      633 GKRYGNSSEIT-TKTLAPRSDNDKEN-----SKSLDPANNESLQOOLEVPHTKEDD 686
Qy      622 VEKNGNLIIPKDHVNIKFAWFDHTYKAPNGYTLLEDLFAITIKY--VEHDERPHSN 679
Db      687 SILANSNIAPEE---LTLVVEANDSSFN---DVTKTFDAYSSFEESLSREHET 737
Qy      680 DG-----WGNASEHVLGKDHSEDPNKNFKADEPVEETPA-----EPEVPOVETE 725
Db      738 DSKPINFISIW-----HKQKQKHQIHVPTKQIIASIQQYKNEQESRVTSD 785
Qy      726 KVE-----AOLKEAVLLAKVT-----DSSLK-----ANA 750
Db      786 KYIIPNAIQFKKKEVNVMSRRVVSFDMDDLNVSGFLPELSDESGFKDLNFANYSNNTNR 845
Qy      751 TETLAGLRNNLTQIOMDN--NSIMAEAEKLLALKSGNSPSSVSEKIN 796
Db      846 PRSFTPLSTKVLNSINIDNPVVEPPEPKSYAEIRNARLSANKAAPN 893

```

RESULT 32

```

US-08-642-846-2
; Sequence 2, Application US/08642846
; Patent No. 5886151
; GENERAL INFORMATION:
; APPLICANT: HOSTETTER, MARGARET K.
; APPLICANT: GALE, CHERYL A.
; APPLICANT: BENDEL, CATHERINE M.
; APPLICANT: TAO, NIAN-JUN
; APPLICANT: KENDRICK, KATHLEEN
; TITLE OF INVENTION: CANDIDA ALBICANS GENE, INTEGRIN-LIKE
; TITLE OF INVENTION: PROTEIN, ANTIBODIES, AND METHODS OF USE
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEETING, RAASCH, GEBHARDT & SCHWAPPACH, P.A.
; CITY: MINNEAPOLIS
; CITY: 119 NORTH FOURTH STREET, SUITE 203
; STATE: MINNESOTA
; COUNTRY: USA
; ZIP: 55401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,846
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: MEETING, ANN M.
; REGISTRATION NUMBER: 33,977
; REFERENCE/DOCKET NUMBER: 110,00280101
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-305-1217
; TELEFAX: 612-305-1228
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1664 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-642-846-2
Query Match      3.2%; Score 133.5; DB 2; Length 1664;

```

Best Local Similarity 17.9%; Pred. No. 0.3;
Matches 159; Conservative 128; Mismatches 364; Indels 237; Gaps 37;

```

Qy      12 VKENNRVSYIDG-----KQATOKENTLTPREVSKREBINAEQIVIKITDQGYTSHSD 64
Db      140 VNEHNAPTYINTSPNKSINKKATPKA--SPKVA-----FTVTNDE 178
Qy      65 HHVYNGKVPYD-----AIISEELMKDPN-YKLKDEDIVNEVGKGVYIKVG 111
Db      179 IHHPDNRVEEEDQOQKEDSVPEPLIQO--WKDPSQVNYDEEDOTNAAVPPRLPHHTK 236
Qy      112 KYVYVILKDAHADNVRTKEINRQKO--EHSQHRGCTPRNDGAVALARSGRY--TTD 166
Db      237 PTFEAL-----LNKNNEVNSPEPALTDMKLKE--NFSNLISLDEKVNLYLSPTN 283
Qy      167 DGYIFNASDI---IEDTGAYIVPHGDHYHI-----PKNELS--ASLAAAEALSLSR 215
Db      284 NNNSKNVSDMDSHLONLQDASKKNTNENIHNLSFALKAKKNDIENPLNLTNADISLRSS 343
Qy      216 GNLSNRTYRPNSDNSTRTNWVPSVSNPGTNTNTNNSNTNSQASQNDISLKKO-- 273
Db      344 GSSQSLQSLRNDNRVLESVPSGPKKVPGL-----SLNDGKGFSDVEVBSLIPRL 396
Qy      274 -LYKLPLSGRH--VESDGLVFDPAQITSRTAGVAVPHGDHYHFIPIYSQSELEERIARI 330
Db      397 SRDKLETTKEHDAPEHNENFIKASTNTNKGQLVSSDHDLDSPDRS-YNHTEQSLNL 455
Qy      331 IPLRYRS---NHWPDSRPEQSPQPTPEPSPGQAPANLKI---DNSLSVSQLVRK 382
Db      456 LNSASQSOISLNALEKQROTOEQTOAEPREETSFSFNIKVKQEPKSNLEFVKTITRK 515
Qy      383 VEGGVYEEBKGISRYVFAKDLPSFTYKNLESLSKQESVSHLTAKKENVAVRDOEFYRK 442
Db      516 -----EPVSATEIAPAPREFSRLIRIKNEDEIAPADIHK 552
Qy      443 AVNLLTEAHKALFNKGRNSDFQALDKLER-LNDESTKEKLVDDLAFAPITHPERL 501
Db      553 KEN---EA-----NSHVEDTDLALKKALNDEESDTJONSTKMSIRPHISDWMK 599
Qy      502 GKPNQIEYTEDEVRIAQLADKYTTSQGYIFDEHDIISDEGAYVTPHMGHSHWIGDSL 561
Db      600 EDSND-----GDRDNDDISFEKSDILNDVSGT-----SDII 632
Qy      562 SDKEKVAQAQYKKEGILPPSPDADYKANFTGDSAAIYNRKGEKRIPLVRLPYWEHT 621
Db      633 GDRYGNSSEIT-TKTLAPRSDNDKEN-----SKSLDPANNESLQOOLEVPHTKEDD 686
Qy      622 VEKNGNLIIPKDHVNIKFAWFDHTYKAPNGYTLLEDLFAITIKY--VEHDERPHSN 679
Db      687 SILANSNIAPEE---LTLVVEANDSSFN---DVTKTFDAYSSFEESLSREHET 737
Qy      680 DG-----WGNASEHVLGKDHSEDPNKNFKADEPVEETPA-----EPEVPOVETE 725
Db      738 DSKPINFISIW-----HKQKQKHQIHVPTKQIIASIQQYKNEQESRVTSD 785
Qy      726 KVE-----AOLKEAVLLAKVT-----DSSLK-----ANA 750
Db      786 KYIIPNAIQFKKKEVNVMSRRVVSFDMDDLNVSGFLPELSDESGFKDLNFANYSNNTNR 845
Qy      751 TETLAGLRNNLTQIOMDN--NSIMAEAEKLLALKSGNSPSSVSEKIN 796
Db      846 PRSFTPLSTKVLNSINIDNPVVEPPEPKSYAEIRNARLSANKAAPN 893

```

RESULT 33

```

US-09-264-604-2
; Sequence 2, Application US/09264604
; Patent No. 6346411
; GENERAL INFORMATION:
; APPLICANT: HOSTETTER, MARGARET K.
; APPLICANT: GALE, CHERYL A.
; APPLICANT: BENDEL, CATHERINE M.
; APPLICANT: TAO, NIAN-JUN

```

```

APPLICANT: KENDRICK, KATHLEEN
TITLE OF INVENTION: CANDIDA ALBICANS GENE, INTEGRIN-LIKE
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESS: MUEITING, RAASCH, GEBHARDT & SCHWAPPACH, P.A.
STREET: 119 NORTH FOURTH STREET, SUITE 203
CITY: MINNEAPOLIS
STATE: MINNESOTA
COUNTRY: USA
ZIP: 55401
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/264,604
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/642,846
FILING DATE: 03-MAY-1996
ATTORNEY/AGENT INFORMATION:
NAME: MUEITING, ANN M.
REGISTRATION NUMBER: 33,977
REFERENCE/DOCKET NUMBER: 110,00280101
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-305-1217
TELEFAX: 612-305-1228
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1664 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-264-604-2

Query Match      3.2%; Score 133.5; DB 4; Length 1664;
Best Local Similarity 17.9%; Pred. No. 0.3;
Matches 159; Conservative 128; Mismatches 364; Indels 237; Gaps 37;

QY 12 VKENRRVSYIOG-----KQATOKTENLTPDEVSKEBSINAEQVIVITDGYVTSBGD 64
DB 140 VNENAPTYINTSPKSIKKAATPKA---SPKVA-----FTVTNPE 178
QY 65 HHVYNGRVYPD-----AIISEELMKDPN-YKLKDEDIVNEVKGVIKYDG 111
DB 179 IHHYDNRVEEDSQOKEDSVEPLIQHQ--WKDPQGNVSDEEDTNASVPTPLHTTK 236
QY 112 KYVYVYLKDAHADNVRTKEEINRQKO--EHSQHBEGGTPRNDGVALARSQRY--TTD 166
DB 237 PTFAPL-----LNKNNEVNSEPEALIDMKLKE--NFSNLSIDEXKNLYLSPTN 283
QY 167 DGYIFNASDI---IEDTDGAYIVPHGDHYHYI-----PKNELS--ASELAAEAFLSGR 215
DB 284 NNNSGKNSVSDMOSHLDQASKNKTNEINHNLSPALKAPKNDIENPLNLTNADISLSS 343
QY 216 GNLSNRTYRRQNSDNTSRTNVPVSNGCTTNTNTSNNSNTNSQASQNSIDISLKO-- 273
DB 344 GSSQSLSGLSRNDNVLESVPSGPKVNPGL-----SLNDGICGFDEVEVESLPLPDL 396
QY 274 -LYKPLPSQRH--VESDGLVFPDPAQITSRTARGAVAPHGDHYHPIPYQMSLELERIARI 330
DB 397 SRDKLETTKEHDAPENHNENFIDAKSTJTNKQQLVSSDDHDSFDRS--YHTEQSILNL 455
QY 331 IPLRRS---NHWWPDSRPREQSPQPTPEPSGQAPANLKI-----DSNLSLVSQLVRK 382
DB 456 LNSASQSOISLALAEKQROTQEOEQTOAAEPEETSFSDNITKVKQEPKPSNLEFVVTYTK 515
QY 383 VGEGVFPEKGISRVVFAKDLPSFTYKNLLESKLSSKQESVSHLTAKKENVAVPRDQEPYDK 442

```

```

DB 516 -----EPVSAETIKAPKREFSSRIIRIKNEDEIAEPADIIHPK 552
QY 443 AYNLTFAHKAFLPNKGRNSDFQALDKLER-LNDESTNKEKVDLLAFIAPITHTBERL 501
DB 553 KEN--BA-----NSHVEDTDALKKALNDDESDTTONSTKMSIRFHIDSWKL 599
QY 502 GKPSQSELYTEDEVRIAQLADKYTTSBGYIFDEHDIIISDEGDVAVTPMHGSHWIGKSL 561
DB 600 EDSND-----GDREDNDISREKSDIINDVSQT-----SDII 632
QY 562 SDKEVAAQAVTKEGILPPSPADVDKAPNTGDSAAAIYNRKVEGKRIPLVRLPYMVEHT 621
DB 633 GDKYGNSSSEIT-TKTLAPRPSDNDKDN-----SKSLDEPANNEISQOQLEVPHTKEDD 686
QY 622 VEYKNGULIIPHKHYNIKRWFEDDHTYKAPNGYITLEDJFATIKYV--VEHPERPSHN 679
DB 687 SILANSSNIAPPEE---LTLPVVEADYSEFN---DVKITDAYISFEESISREHET 737
QY 680 DG-----WGNASEHVLAGKDHSEDPNNKFKADEEPVEETPA-----EPEVPOVETE 725
DB 738 DSKRINFISTW-----HKQEKQKHQIHKVPTKQIIASYYQYKNEQESRVTSD 785
QY 726 KVE-----AQLKEAVLLAKVT-----DSSLK-----ANA 750
DB 786 KVKIPNAIQFKKFKXVVMRSRRVSPMDDLNVSQFLPELSEDSGFXDLNFAVNSNNTNR 845
QY 751 TETIAGLRNNITLQIMDN--NSIMAEBEKLALLKGSNPPSVSXEKIN 796
DB 846 PRSFTPLSTKNVLSINDNPVVPPEPKSYAEIRIARLRSANKAAPN 893

RESULT 34
5268270-2
; Patent No. 5268270
; APPLICANT: Meyer, Thomas F.; Halter, Roman; Pohner, Johannes
; TITLE OF INVENTION: PROCESS FOR PRODUCING PROTEINS USING GRAM
; NEGATIVE HOST CELLS
; NUMBER OF SEQUENCES: 6
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/171,872
; FILING DATE: 01-JUL-1987
; SEQ ID NO: 2
; LENGTH: 1507

Query Match      3.2%; Score 132.5; DB 6; Length 1507;
Best Local Similarity 18.1%; Pred. No. 0.31;
Matches 183; Conservative 125; Mismatches 305; Indels 399; Gaps 53;

QY 53 ITDQGYVTSHG-----DHYHYNGKVVPYDAIISEELMKDNPYKLDKEDIYNEVKG 104
DB 523 ITNNSLSVHSIQNDYDEDDSYTPRRPRPQ--GKDLVYKARY-----EGG 566
QY 105 YVIRVDGKYVYVYLKDAHADN-----VRTKEEINRQKQESHQR--EGG-----TP 148
DB 567 YALSGGRILNAPMENCVAENMDWI PMGYTQEARKQAMNHKNNRRI GDEGFPDEENGK 626
QY 149 RNDGVAL---ASQGRYTTDDGYITNASDIIEDTDGAYI---VPHGDHY----- 192
DB 627 GHNGLALNINFGKSAQNRFLITGALNKG--ISVTOGNVLISGRPTPHARDFVKKSSARK 685
QY 193 --HYIPKNEU-----SASELAA--AEAFLSGRG-----NLSNSRTYR 225
DB 686 DAHFSKNNEVVFEDDMINRTFKAAEIAVNSQASFSGGRANDDTANITATIDNAKVNIGYK 745
QY 226 RQNSDNTS-RTNWVPSVS-NPGTTN-----TNTSNNSNTNSQAS----- 262
DB 746 --NDEVCVRSDDYGYTCTGNSLSDKALNSFDKATKINGVNLNQNAAALVLGKAAALMGK 803
QY 263 ---QSNIDISLKL-----QLYKPLPSQRHV----- 284
DB 804 IQGQNSGRVSLNQS KWHLTGDSOVNHLISLADSHIHNLNNSADAOSANKYHTIKINHLSGN 863

```

```

QY 285 -----BSDGI-VFDPQITSRITARGVA 305
DB 864 GHFHYLTDLAKLGDVYLKVEASAGHYOLHVQNKTEGPNOEGDLDFDASSVQDRSRLFVS 923
QY 306 VHHGHVHIFPYSQMS---ELEERIRARIIPLRYSNHWVPDSPEPSPQPT-----354
DB 924 LAM---HYVDLGALEKRYTITKTEGTRLY-----PRVANGNRPVAPSPAPANTASQAO 973
QY 355 -----PE-----PSPQOPAPNLIKIDSSSLV-----SOLV 380
DB 974 KATQDGAQIAKQNIIVNAPSPQAOAEIRQAOAAEOYKQOAAEAKEVNAQKEEAK 1033
QY 381 RYVGE-GYVFEKGISRYVPKADLPSETVKNIESKLSQESVSHLTAK---KENVAP 434
DB 1034 RKAAEIARQOEERKAAELAAQOKAERKARELARQKAEASHQANAKPKRRRRALIP 1093
QY 435 RD-----OEFDKAVNLTGAH 451
DB 1094 RPPAPVSLDDYDAKDNSESSIGNLARVPIRMRELINDYEEIPLLELDEAEERROAT 1153
QY 452 KALFXNKGNSDFQALDKLERLNDSTNKEKLVDDLLAFIATPHERLKGNSQIEYT 511
DB 1154 Q--FHKSXNR--RAISS--EPSSDDASESVSTSD-----KHPDNTLHEKVEYTA 1199
QY 512 EDEVRIAQ-----LADKYYTSDGYIFDE---HDIISDEGDAYVTPHMGSHWIGXD 559
DB 1200 GQOPRAOQPTQAAQADAVSTWNSALSDAMASTOSILLDTG-ALUTH-----IAQK 1252
QY 560 SISDEKVAQAATYKKEKILPSPDADYKAMPFG---DSAAAIYNKVEKGRIPVRLPY 616
DB 1253 SRADEAK-----NSVMMENTGYRDYASAOYRFSKRTQTOGIDIR 1294
QY 617 WHEHTVEKXGNLIIPKHVHNKIPAMPFDHTYKAPNGYLTLEDLPAITIKYVEHEHDERP 676
DB 1295 SISEMJOIG--VLYTSDSQHT---PDQAGK--NTFVQANLIG--KYL-----1335
QY 677 HSNDBGKNSAEHLGK-KDHSDDPNK-NFKADEPVEETPAPEVPQVETEKVE-----728
DB 1336 --NDAYVAGDIGAGSLRSRLQTOQKAF-----NRTSIOGTGLTGMLTKINGEIEVP 1386
QY 729 -----AOLKEAEVLLAKYTDSLSKAN--ATEYL-AGLR-----NNLTLO 764
DB 1387 SAGIRYSRLSSADY---KLGDSSVKVSMVAVKTLTAGLDPAVRFKGNLTVK 1435

RESULT 35
US-09-457-708-2
/ Sequence 2, Application US/09457708
/ Patent No. 6326483
/ GENERAL INFORMATION:
/ APPLICANT: Kwiatkowski, David J.
/ APPLICANT: Sampson, Julian R.
/ APPLICANT: Povey, Sue
/ APPLICANT: van Slegtenhorst, Marjon
/ APPLICANT: Halley, Dicky
/ TITLE OF INVENTION: Compositions and Methods Based on
/ TITLE OF INVENTION: Sclerosis-1 (TSC-1) Gene and Gene Product
/ NUMBER OF SEQUENCES: 28
/ CORRESPONDENCE ADDRESS:
/ ADDRESS: Vinson & Elkins
/ STREET: 1455 Pennsylvania Avenue, N.W.
/ CITY: Washington
/ STATE: D.C.
/ COUNTRY: U.S.
/ ZIP: 20004-1008
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/457,708
/ FILING DATE:

```

```

/ CLASSIFICATION:
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Sanzo, Michael A.
/ REGISTRATION NUMBER: 36,912
/ REFERENCE/DOCKET NUMBER: BR1331/42002
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (202) 639-6585
/ TELEFAX: (202) 639-6604
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1164 amino acids
/ TYPE: amino acid
/ STRANDEDNESS: not relevant
/ TOPOLOGY: not relevant
/ MOLECULE TYPE: protein
/ HYPOTHEICAL: NO
/ ANTI-SENSE: NO
/ US-09-457-708-2

```

```

Query Match 3.2%: Score 132; DB 4; Length 1164;
Best Local Similarity 18.8%: Pred No 0.22;
Matches 155; Conservative 122; Mismatches 315; Indels 234; Gaps 35;

```

```

QY 102 KGGYIKYDGYVYLYKQAAHA-----DNVTRKEIRNRQKQ 137
DB 166 KGVAAEV---YLVHLHSAVVALFRLYGMYPNCFVSLRSHYSKMNLETFEEVYKPMW 224
QY 138 EHSQ-HREGGTPRNDGAVVALASQGRYTTDDGYIRNASDIIEDT---GDAYIVPHGHY 192
DB 225 EHVRIHPELVYTSKHELD-PRRWKRELETHDVICAKISLPTPEASVEDGYSVSH---279
QY 193 HYIPKNEISAS-ELAAEAFLSGRNLNSRIRYRQNSDNTSRITMWVPSVNG--TTNTN 250
DB 280 -----QISARPHNSADVTTSFYADTONSYCATSTPYSTSR---LMLNMPGQLPOTL 330
QY 251 TSNNSNTNSQASQSN-DIDSLKQLYKPLSGRHVESDGLVDPDAQITSRITARGVAP--307
DB 331 SSPSTRLIPEPQATLWSPSWVCGMTTPTSPGNVPPD-LSHPYGKVFPTTAGKGTPLG 389
QY 308 -----HGHHYHIFPYSQMS---ELEERIRARIIPLRYSNHWVPDSPEPSP 351
DB 390 TPATSPAPALCHSDDYVAHISLPQATVTPPRKEEBMDSARPLRQHILNDRGSEEP--447
QY 352 QTPPEPSPQOPAPNLIKIDSSSLVSQVRYKVGEGVFEKGISRYVFAKDLPE--TVK 409
DB 448 -----PESKGSVTL-----SDLPG-----FUGDLASEEDSIE 474
QY 410 NLESKLSQESVSHLTIAKENVAPR---DQEFY-DKAVNLTTEAHKALFXNKGNSDQ 465
DB 475 KKEBAALSRLESLITTAEAEPVVRGGRFDSFYRDSLPGSQKTHSAASSQGSQSVNPE 534
QY 466 ALDKLIERLNDSTNKEKLVDDL-----LAFIAP---ITPHERL 501
DB 535 PHLSSLDKLGPTPQOAFPTIDLPGSADESPAGRECQTSLETSIFTPSPCKIPPTFRV 594
QY 502 GKNPQOIEYTHD--VYRIAQADKYYTSDGYIFDEHDIISDGDAYVTPHMGSHWIGXD 559
DB 595 GFGSQOPPEYDHLFEVALPKTA-----HHFVIRTEBELKKAKANTBEDGPV 641
QY 560 SISDEK---KVAQAATYKKEGILP--PSPDAD---VKANPTGSAAN-----598
DB 642 STSPWEVLDRLIQGADAHASKELNKLPLPSKVDWTHPGSGSPSBEIRLRLQOULLHNO 701
QY 599 -IYNRVKGEKRI-----PLVRLPYWHEHTVEKXGNLIIPKHVHNKIPAMPFDHT 649
DB 702 LLYERFKQOHALRNRRLRKVIAKAAAEHNAAMKD-----QLKLOEKDIOM 749
QY 650 YKAPNGYLTLEDLPAITIKYVEHPD-----ERPHSNDKGNASAHVGLKGDHSD 698
DB 750 WTV---STQKEQAYYNLOQEQDRTWVTKLHSQIRLOLQDREEFYNGSOELQTKLEDCC--803
QY 699 PKNFKADEEPVEETPAPEVPQVETEKVEAOLKEAEVLLAKYVTSLSKANATE-----752

```

Db 804 --RNMIAE-----LRLELKANNKVCHEITELLQSOKLSNSESVOOOMEFL 848

Qy 753 -----TLACLRNNLTQIMDNNSIMAEKLLALLKGSNPSSVSXK 794

Db 849 NRQLLVGEVNELYLEQJONKH--SDTTKEVEMKAAVRKELEKNR 892

RESULT 36

US-09-950-046A-2

Sequence 2, Application US/09950046A

Patent No. 6548258

GENERAL INFORMATION:

APPLICANT: Kwiatkowski, David J.

Sampson, Julian R.

Povey, Sue

van Stegenhoret, Marion

Halley, Dicky

TITLE OF INVENTION: Compositions and Methods Based Upon the Tuberculous Sclerosis-1 (TSC-1) Gene and Gene Product

NUMBER OF SEQUENCES: 28

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Vinson & Elkins

STREET: 1455 Pennsylvania Avenue, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: U.S.

ZIP: 20004-1008

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/950,046A

FILING DATE: 12-Sep-2001

CLASSIFICATION: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Sanzo, Michael A.

REGISTRATION NUMBER: 36,912

REFERENCE/DOCKET NUMBER: BR1331/42002

TELECOMMUNICATION INFORMATION:

TELEPHONE: (202) 639-6585

TELEFAX: (202) 639-6604

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 1164 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: not relevant

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-950-046A-2

Query Match 3.2%; Score 132; DB 4; Length 1164;

Best Local Similarity 18.8%; Pred. No. 0.22; Mismatches 315; Indels 234; Gaps 35;

Matches 155; Conservative 122;

Qy 102 KGGYVYKDGKYYVYLKDAHA-----DNVTKKEINRQK 137

Db 168 KPGVAEV---YLVLHLSVYALFRLYGMYPQNVFSLRSHYSKMELEFEEVYKRM 224

Qy 138 EHSQ-HREGGTPRNDGAVALAASQGRYTTDGYIFNASDIIEDT---GDAYIVPHGDY 192

Db 225 EHVRIHPLVLTVGSKDHELD-PRRWKRLFTHDVICAKISLDPTEASYEDGYSVSH--- 279

Qy 193 HYIPKNEISAS-ELAAAFALSGRNLNSRTYRQNSDNTSRRTWVSVSIPG--TTVTN 250

Db 280 -----QISARPPHRSADVTTSPYADTONSYGCATSTPYSTR---LMLLNNPQLPQTL 330

Qy 251 "TSNNSTNSQASQSN-DIDSLLKOLYKPLSGRHVESDGLVDPQAQITSRTARGAIVP-- 307

Db 331 SSPSTRLLTEBPQATLWSPSNVCMGTTPTSPGNVPPD-LSHPYSKVFGTTAGKGPPLG 389

Qy 308 -----HGDHFIPIYSQMS-----ELEERLARIIPLRYSRNMHWPDSPPEOSP 351

Db 390 TPATSPPPALCHSDDDVHISLPQATVTPPKKEERMSARBCLRQHLNLDROGSEB- 447

Qy 352 QPTPEPSPGPQAPADNLKIDSNSLSVSQLVRKVGEGYVEEKGISRYVAKDLPSE--TVK 409

Db 448 -----PGSKGSVTL-----SDLPG-----FLGDLASEBDSIE 474

Qy 410 NLESKLKQESVSHTLTAKKENVAPR--DOEFY-DYAVNLTAAHAKALFNKGRNSDFQ 465

Db 475 KDEEAAISRLEISITTAEAEPVPRGCFDPSFYRDSLPGSQRTKTHSAASSOGASVNP 534

Qy 466 ALDKLERLNDESINKELVDDL-----LAPLAP-----TTPERL 501

Db 535 PLHSSLDLGPDTYKQAFPTIDLPQGSADSPADGRCQTSLETSITPSPCKLPPTRV 594

Qy 502 GKPSQLEYTED--EVRILAQADKYTTSDGYIPDEHDIISDEGDAYVTPHMGSHWIGKD 559

Db 595 GFGSGQPPYDHLFEVALPKTA-----HHFVIRKTEELLKAKGNTEEDGVP 641

Qy 560 SLSDKE-----KYAAQAYTKEGILP-PSPAD--VKANPTGDSAAA----- 598

Db 642 STSPMEVLDRLIQGADASHKELNKLPLPSKSVDMTHFGSPSPDEIRTLRLDQLLLHMQ 701

Qy 599 -IYRVKGEKRI-----PLVRLPYMVEVTEVVKXGNLIIPIKDHYNIKFAMPDHT 649

Db 702 LLYERFROQIALNRRLLRKRVIAALAEHNAAMKQ-----QKLOEKDIQM 749

Qy 650 YKANGYTLEDLFAITIKYVEHPD-----ERPHSNDGNGNASEHYLGKXDSED 698

Db 750 WKV-----SLOKEQARVYNOLOEQRDWTYKLSQIRQLOHDEEYVNSQELQTLDEC-- 803

Qy 699 PNKQFKADEEPVEETPAEPVPOVETEKVEAQLEAEVLLAKTDDSSKANATE----- 752

Db 804 --RNMIAE-----LRLELKANNKVCHEITELLQSOKLSNSESVOOOMEFL 848

Qy 753 -----TLACLRNNLTQIMDNNSIMAEKLLALLKGSNPSSVSXK 794

Db 849 NRQLLVGEVNELYLEQJONKH--SDTTKEVEMKAAVRKELEKNR 892

RESULT 37

US-09-071-035-458

Sequence 458, Application US/09071035

Patent No. 6448043

GENERAL INFORMATION:

APPLICANT: Gil H. Choi

NUMBER OF SEQUENCES: 496

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/071,035

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: A. Anders Brookes

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB369P2
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 458:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2032 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-071-035-458

Query Match 3.2%; Score 132; DB 4; Length 2032;
 Best Local Similarity 19.5%; Pred. No. 0.54;
 Matches 185; Conservative 127; Mismatches 316; Indels 320; Gaps 51;

```

QY 33 LTPDEVSKREGINAEQIYI---KITDQ---GYTSHGD---HYHYNGKVPYD-AII 79
DB 483 LTSD--NNPENGDAEATVATYGMKDKENIDYDEANQFTWEINYNVGEQTI PKDQAVI 540
QY 80 SEEL---LMKDPN---YKLKDEDIVNEVKG---YVIKDGKYYV-YLKDAH 122
DB 541 TDTMGDNLTFEPDSLHLYSVTFDDKGNVGAELVEGDKYKVVINGDSFAIDFLHDVTG 600
QY 123 ADNVRTKEINRQKQHSQHREGGTRPNDGAV--ALARSQGRYTDDGYIFNASDIIEBT 180
DB 601 AKIDYKTKVD-----GIVEGDVAVNNRVDTGQHSDDG-TASQONIIKXT 647
QY 181 GDAYIVPHGDHYHYIPKNELSASELAABAFISGRGNLSNRTYRQNSDNTSRTNWVPS 240
DB 648 GAV-----DY-----QNSTIGWTLAV-----NONNYLMENAVITDYEPPVG 684
QY 241 VSNPGTNTNNTSNNTSQAQSNIDISLKLQYKPLSQHVESDGLV-FDPAQITSR 299
DB 665 LTM--VPSNLVVKDITTGALTLGKD-----FWEIETR--NADGETGKVSFI--- 728
QY 300 TARGVAVPHGDHYHYIPYQSELEERLARIIP-L-RYR-----SNHWVDSNP 346
DB 729 ---GAVAKTSDAFH-ITYTTFDVTIELDANNRDLHYRTAAIDWDEAGNNHSEDSRP 784
QY 347 EOPSPQ-PTPEPSPGPQAPNLIKID-----SNSLSVQLVR-----KVG 384
DB 785 FKPLPAFLDIAQKSGVNAVTKETITWTLAVNLSNNRLVDAFLDITLTQYTIAGSLKXY 844
QY 385 EGYVPEEKISRYVPAKDL-----PSETVKN-----LESKLSKQESVSH 423
DB 845 EENTKPDGSEVEKVKPTQPLDITMEBPSEKQNTWRVDPNDSTRYVIEFKTSVDEKYLE 904
QY 424 TLTAKENAVPRDQEFYDKAVNULTFAHKALEFANKGRNSDFOALDKLERLNDSTNK-- 481
DB 905 GSAS-----YD-----NTASTNQGSSND--VTGKVASIQHGSEYVKKG 941
QY 482 -----EKLVDLLAFLABITPHERLKGKNSQIETEDEVRIAQLAD 522
DB 942 EYHKDPRDHVVMWINGASVLDVLTDTF--SPNQVLDESLVITYTN----- 990
QY 523 KTTSDGYITFDEHDIISDEGDAYVTPRMGSHWIGKDSLSKEKVA-----AQAYTKE 575
DB 991 --VTEDGTTTPKSYILEGKDYTLE-----VTTDNETGQOKIVVMAHIEAPYMEY 1041
QY 576 KGIILPSP--DADVANP--TGDSAAALYNVKGKRIPL-----VRLP 615
DB 1042 RSLVSSAAGSTDTYVSNQVSTINGSGSEVAVGDDNGVVVDIDHSGHATGTGKIQLKKT 1101
QY 616 YWVEHTV-----EYKNGNLLIIPKDHYNIKFAWFDHTY-----KAPNGYTLLED 660
DB 1102 AMDETTILAGAHFQIWDQAKTQVLRGCTVATGVTIFGGLPGQYIYLVETAKAPGTYVSD 1161
QY 661 LFA-----TI-----KYYVEHDERPHSNDGKGNASEHVLGKQHS 696
DB 1162 ELAKGRVITIDETSABGAQPTIINKDVNVKFLKMDK-----GKK--- 1203

```

```

QY 697 EDPNKNFPADEEPPV-----EETPAEPEVP---QVETEKVBA---QLKAEVLLAKYVD 743
DB 1204 -LVNARFLEHNAVTPFTTHWESEVPLAPRTNANGQLEVDLSLKGLYQFTIEIAPFGYLLD 1262
QY 744 SLKA-NATEITLAGLRNNLTQIOMDINSIMAEKTLALLK--GSPN 787
DB 1263 TTPKRFIVTQNTSG-----QIRDVHVKMVNYQSAELIKMDQAGNP 1303

```

RESULT 38

US-09-071-035-462
 ; Sequence 462, Application US/09071035
 ; Patent No. 6448043

GENERAL INFORMATION:

APPLICANT: Gil H. Choi
 TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
 NUMBER OF SEQUENCES: 496
 CORRESPONDENCE ADDRESS:

ADDRESSER: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage

COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/071,035
 FILING DATE:

CLASSIFICATION:
 PRIOR APPLICATION DATA:

ATTORNEY/AGENT INFORMATION:
 FILING DATE:

NAME: A. Anders Brooks
 REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB369P2
 TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 462:
 SEQUENCE CHARACTERISTICS:

LENGTH: 2032 amino acids
 TYPE: amino acid

STRANDEDNESS: single
 TOPOLOGY: linear

MOLECULE TYPE: protein
 US-09-071-035-462

Query Match 3.2%; Score 132; DB 4; Length 2032;
 Best Local Similarity 19.5%; Pred. No. 0.54;
 Matches 185; Conservative 127; Mismatches 316; Indels 320; Gaps 51;

```

QY 33 LTPDEVSKREGINAEQIYI---KITDQ---GYTSHGD---HYHYNGKVPYD-AII 79
DB 483 LTSD--NNPENGDAEATVATYGMKDKENIDYDEANQFTWEINYNVGEQTI PKDQAVI 540
QY 80 SEEL---LMKDPN---YKLKDEDIVNEVKG---YVIKDGKYYV-YLKDAH 122
DB 541 TDTMGDNLTFEPDSLHLYSVTFDDKGNVGAELVEGDKYKVVINGDSFAIDFLHDVTG 600
QY 123 ADNVRTKEINRQKQHSQHREGGTRPNDGAV--ALARSQGRYTDDGYIFNASDIIEBT 180
DB 601 AKIDYKTKVD-----GIVEGDVAVNNRVDTGQHSDDG-TASQONIIKXT 647
QY 181 GDAYIVPHGDHYHYIPKNELSASELAABAFISGRGNLSNRTYRQNSDNTSRTNWVPS 240
DB 648 GAV-----DY-----QNSTIGWTLAV-----NONNYLMENAVITDYEPPVG 684
QY 241 VSNPGTNTNNTSNNTSQAQSNIDISLKLQYKPLSQHVESDGLV-FDPAQITSR 299

```


QY 661 LFA-----TI-----KYYVEHPDERPHSNDGWNASEHVLGKKDHS 696
 Db 1162 ELAKGAVITIDETSAGAQPTIIKNDVKNVFLKXDEK-----GKK--- 1203
 QY 697 EDNKNKFKADEEV-----EETPAEPEV-----QVETKEVA---OLKEAEVLLAKVTD 743
 Db 1204 -LVNAFKEKLEHAVTFTFTHWEVPLAFDRFNANGOLEVDSLKPGLYQFTIEAPTGYLTD 1262
 QY 744 SSLKA-NATETLAGLRNNLTLQIMDNNSIYAEAEKLLALLK---GSP 787
 Db 1263 TTPKRTIVQNTSG-----QIRDVHVKNLNYQGSALITKQDAQANP 1303

RESULT 40
 PCT-US93-03077-1
 ; Sequence 1, Application PC/TUS9303077
 ; GENERAL INFORMATION:
 ; APPLICANT: Board of Regents, The University of Texas System
 ; APPLICANT: Gaylor, Richard B.
 ; APPLICANT: Wu, Foon Kin
 ; TITLE OF INVENTION: PROTEIN CELLULAR FACTOR USEFUL FOR
 ; TITLE OF INVENTION: REGULATING GENE EXPRESSION
 ; NUMBER OF SEQUENCES: 7
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Arnold White & Durkee
 ; STREET: P.O. Box 4433
 ; CITY: Houston
 ; STATE: Texas
 ; COUNTRY: USA
 ; ZIP: 77210

COMPUTER READABLE FORM:
 ; MEDIUM TYPE: floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent in Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: PCT/US93/03077
 ; FILING DATE: 19930331

CLASSIFICATION:
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/862,025
 ; FILING DATE: April 2, 1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Kammerer, Patricia A.
 ; REGISTRATION NUMBER: 29,775
 ; REFERENCE/DOCKET NUMBER: UFD270PCT
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 713-787-1540
 ; TELEFAX: 713-749-2679

TELEX:
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1093 amino acids
 ; TYPE: AMINO ACID
 ; STRANDEDNESS: unknown
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 ; PCT-US93-03077-1

Query Match: 3.1%; Score 131; DB 5; Length 1093;
 Best local similarity 16.7%; Pred. No. 0.24;
 Matches 141; Conservative 129; Mismatches 356; Indels 216; Gaps 24;

QY 75 YDAIIEELMKDPNKLKDEDIVNEVGGYVVKDGKYYVYLKDAHAD---NVRTKE 130
 Db 177 HEETVKNKESDMKVPVSLKVSSEVIDVK-----TTMESISNTSTQSLTAETKD 224
 QY 131 EIRKQOEHSGHEGTPRNDGAVALARSGRTTDDGIVFNASDIIEDG----- 181
 Db 225 IALEPKQKHEDQSNTP---SPVSTFSSGTSTSDIEVLDEHVSISESSASSROETTD 281
 QY 182 -----DAYIVPHGDHYHYPKNELSASELAALAFISGRNLSNSRTYR 226

Db 282 SKSLHLMQTSFOLLSASACPEYNRLDQFKLTESCCSSDAFERIDFSVQSLDSRSVSE 341
 QY 227 QNSDNTSRNWNPEVSNPCTTNTNNSNTNSQASQNSDIDSLKQLYKLPLSQNHVES 286
 Db 342 INSDDELSKGYALV---PIIVNSSTPKSTVESAECKSEEVNTL---VITFEAEWER 395
 QY 287 DGLVFPDAITSRTAGVAVPPGDHYFI-----PVSQNSE----- 322
 Db 396 SGRSATPVNCEQPDILVSTPINEGQTVLDKVAEOCEPESQPEALSEKEDCKTVEFLN 455
 QY 323 --LEERIRAIIPLRKYSNHWVDPDSRPEQSPQPTPEPSPQOPAPNLKID-----S 371
 Db 456 EKLEKREAQLSL-----SKKALLEAFNLDXDMFVRKESSS 495

QY 372 NSLSVQLVRKYGEGVEFEKGISRYVPAKDLPSFTVKULBSLSQESVSHLTAKKEN 431
 Db 496 ISSLKDEFTQRIAE---EKVQOLACKERDAKEIKYIKELATRLNSETADLLKKE 551
 QY 432 VAPRQEFYDKAVNLTTEAHKALFNKGRNSD---FOALDKLBERLNDSTKERTVLD 487
 Db 552 -----DEQIRGLMEGCKLSKQQLHNSNITKLRKADKENEMVAKLNKKVLELE 602
 QY 488 LLAFLAPI-----THPERLGRNSQIEYTED-----EYRIAQLADKYTTSDGYIPD 533
 Db 603 ELQHLKQVLDGKEVEKQRENIKLNMSVBERQEKOLGLQVMDLEEKNSIQAAADS 662
 QY 534 EHDIISEGDGAYTPMGMHSHWIGKDSLDEKVAQAATYKKGILPPSPDADVKNAPPG 533
 Db 663 AYKELTDL-----HKANAADSEAOEALSHREMAKEE----- 695

QY 594 DSAATVYNVKGKRIPLVRLPYMVEHTVEVNGNLI-----PHKDHYNIKPAM 644
 Db 696 --LSAALKAQOEBAE-----QOQETLAIQGDRLRLQRTQQAARKEDYLREITEG 745

QY 645 PDHTYKAPNG-----TYLEDLFAIKYVHPDE-RPHSNDGWNASE 687
 Db 746 LQRLQEAENRQELQSQSVSTTRPLLRQIENLQATLSQTSWELEKNLSRLEQSG 805

QY 688 HVLGKKDSEDPNKNFKADEBEVEETPAPEVPQVETEVEAOQLKAEVLLAKVTD--- 743
 Db 806 LAAAVERRAATEELANKIQMSMSQNSLIRQENSFOQLSEKURLCLDEENNR 865

QY 744 -----SLKANATETLAGLRNNLTL--QIMDNNSIYAEAEKLLALL-----KGSNPS 789
 Db 866 YQVELENLDEYVITLETRKEKTLNLSQLEMERMKVEQERKKAIFTQETIKERKPPS 925

QY 790 VS 791
 Db 926 VS 927

RESULT 41
 US-09-417-197-49
 ; Sequence 49, Application US/09417197
 ; Patent No. 6518021

GENERAL INFORMATION:
 ; APPLICANT: Ole THASTRUP, et al.
 ; TITLE OF INVENTION: A Method For Extracting Quantitative Information Relating To An
 ; FILE REFERENCE: 3759-0110P
 ; CURRENT APPLICATION NUMBER: US/09/417,197
 ; CURRENT FILING DATE: 1999-10-07
 ; NUMBER OF SEQ ID NOS: 143
 ; SOFTWARE: Patent in version 3.0
 ; SEQ ID NO 49
 ; LENGTH: 968
 ; TYPE: PRT

ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: EGFP-p85alpha fusion
 US-09-417-197-49

```

Query Match      3.1%; Score 130.5; DB 4; Length 968;
Best Local Similarity 17.0%; Pred. No. 0.22;
Matches 146; Conservative 105; Mismatches 253; Indels 353; Gaps 41;

OY 64 DHY-----HYNGKVPYDAITSEELMKDPYKJLKEDIVNE--VKGGY 105
DB 181 DHYQONTPIGDPVLLPBNHY-----LSTQALSKDPNEK--RDHVVLEFETAAGI 230
OY 106 VIKVQKGVVYLKD--AAHADNVRTEKEINROKQESQREGE--TPRNDGAVALARSGR 162
DB 231 TLGMELYKSGLRMSAEGYQYRALYDYKKEREEDIDLHLGILTVNKGSLVALGSPDQ 290
OY 163 YTTDD--GYFNASDIIEDTDGYIVPHGDHYHYPKNELSS-----AS 203
DB 291 EARPEIGMLNGYNVTGGERDF--PQTYVEYIGRKKISPTPKPRPRPLVAPGSS 346
OY 204 ELAA-----AEAF-----LSGRGLNSRTYRRONSNTSRT 235
DB 347 KTEADVEOQALTPDLAEQFAPPDIAAPLLIKLVEAIEKKG--LECSTLYRTQSSNLAEI 405
OY 236 NWVPESVSPGTTNTSNTNSNSQASQSDNI-----DSLKLQYLK--PLSGRHVESDGL 289
DB 406 ROLDLCTP-----SVLEMTDVHVLADAFKRYLLDLPNVPVIPAAYSEMI 451
OY 290 VFDP-----AQITSTARGAVAPHGDHY-----HFIPYSQMSLEERLARI 331
DB 452 SLAPVOSSEYIQLKKLIRSPILPH--QYVLTLOYLKHFPLKSQTSKULMAARVL 509
OY 332 -----PLRYR-----SNHWVDSRPEQSPQPTPEPSFGOPAPNL 367
DB 510 EIFSMLRFRFMASSDNTENLIKIVIELISTEW-----NERQPA PALPAP--PKPTVANN 564
OY 368 KIDSSSL-----VSQLVKRVGEG----- 386
DB 565 GNNNNMSLOANAEWYMGDISREEVNKKLDYADGTFLVBDASTKMGDYTLTKRGCKNNKL 624
OY 387 -YVFEKGISRYVFAKDLPSETVKNL-----ESKLSK 417
DB 625 IKIFHRDG--KYGFDPLTFSSVVELINHYRNESLAQNPVKDVLVYPSKYQDDQYVK 682
OY 418 QESVSHLLTAKKENVAPRDQEF-----YDKAVNLLTEAHKAL----- 454
DB 683 EDNI-----EAVGKKLHEVNTQFOEKSREYDRLYEETYSQEIQWKRTAIEAFNET 734
OY 455 -----FXNKGKNSDPQAL-----DKLLERLNDSTNKEKVLDDL 489
DB 735 IKIFEEQOQOTORYSKEYIEKFKREGNEKEIQRIIMHNYDKLSRSEIIDSRRLEEDL- 793
OY 490 AFLAPITHPERLKGKNSQIEYTEDEVRI-----AQLADKYTTSQGYIF----- 532
DB 794 -----KKQAAEYREIDKRNNSIKRDLIQLRKTRDQYLMMLTQKGVQKRLN 839
OY 533 -----DEHDIISDEGDAVYTPHMGSHW-IGKDSLSPKEKVA-----AQAYTK 574
DB 840 EWLGNENTEDQYSLVEDEDED--LPHHDEKTNVSSNRNKAENLLRGKRDGTFLVRESS 896
OY 575 EKGILPES--PDADVK-----ANPFGDGAALYNNVKGKRIPLVRLYLMYEHYEVANG 627
DB 897 KQGCYACGVVVDGVEKVCVINKTARGYGAPEYNLYSSLKELVL-----HYOHT----- 945
OY 628 NLIIIPHKDHYHNKIFAW 644
DB 946 -SLVQHNDSL-NVTILAY 960

```

```

; APPLICANT: Horrocks, U. Suzanne Hoyle
; APPLICANT: Hough, Christine
; TITLE OF INVENTION: Canine Factor VIII Gene, Protein and Methods of Use
; FILE REFERENCE: 1669.0010002/JAG/BJD
; CURRENT APPLICATION NUMBER: US/09/324,867A
; EARLIER FILING DATE: 1999-06-03
; EARLIER APPLICATION NUMBER: 09/035,141
; EARLIER FILING DATE: 1998-03-059
; EARLIER APPLICATION NUMBER: 60/039,953
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 2115
; TYPE: PRT
; ORGANISM: Sus spp.
US-09-324-867-5

Query Match      3.1%; Score 130; DB 3; Length 2115;
Best Local Similarity 19.7%; Pred. No. 0.82;
Matches 141; Conservative 84; Mismatches 233; Indels 258; Gaps 37;

OY 149 RNDGAVALARSGRYTTDDGYIFNASDIIEDTDGYIVPHGDHYHYPKNELASASELAA 208
DB 699 RNRGWTALLKV-----YSCD-----RDIGDY-----DNTYEDIP----- 728
OY 209 EAPLSGRGNLSNRTYKRONSDNTRTNWVPSVSNPGTTNTSNTNSQASQSDNID 268
DB 729 -GFLSGKNVLEPPSPFONSPSPASOKOFOTITSP--EDVELDPQSGERTQALAEUSVP 786
OY 269 SLKLQYLKPLSQHVESDGLVF--DPAQITSTARGAVAPHGDHYHYPYSQMSLEEE 325
DB 787 S-----CDGSMILGQNP-----PHG-----SSSDLOE 810
OY 326 RIARIIPKRYKSNHWVDSRPEQSPQPTPEPS-----GPOPAPNL-KIDSNS 374
DB 811 -----ARNADADDYLPGARERNTPAASAAARLAPLHLSAERVLTPPEKEKELKIDSXMS 863
OY 375 LVSQLVKKVGGYVFEKGISRYVFAKDLPSETVKNLKESKQESVSHLLTAKKENVAP 434
DB 864 SSSDLKLT-----SPTTSDT-----LSAERTHSIGPPHOVNF 899
OY 435 RDQ-----EFYDKAVNL--LTEAHKALFXNKGKNSDPQALDKLERLNDSTNK 481
DB 900 RSQIGATVLCGNSSHFIAGVPLGSTEDHES--SLGENVSPVSDGIEFK--ERAHGP 954
OY 482 EKLVDLLAFLAPITHPERLKGKNSQIEYTEDEVRI-----AQLADKYTTS 527
DB 955 ASLTRKDVLPFKVNIS--LVKTNKARVYLTNRKIHIDDAALLTENRASATFMDKNTTA 1010
OY 528 DGYIFDEHDIISDEGDAVYTPHMGH--SHWI-----GXDSLSPKE----- 565
DB 1011 SG-----LNHVSNNWIKGPLKPNLSEGRPSPELLTSSGSGSVK 1050
OY 566 -----KYAAQAYTKEG--ILPPSP-----DADYKANPTGSAALAYNRVKGK 607
DB 1051 GSSGCGGRIKAYVBEELSKKEMMLPNSLFTLTNSADVQGNLT-----HSQCK 1101
OY 608 -RIPLVRLPYVNEHTVEKGNLLIIPKDHYNIKFAMFDDHYTKAPNGYTLDELPAATK 666
DB 1102 SREMERERKLVQEKVDLPQYVATGTGNFLRNI-----FHQSTPVSVEGFGDS----- 1151
OY 667 YVVEHPDERPHSNQKGNASHEVVGKDHSDPRKNFKA--DEEPVETPAPEVP----- 720
DB 1152 -HAPVPQDSRSLNDSABRAETHI-----AHFSALIREAPL--EAPNRTGPGPRS 1198
OY 721 OVETEKYEALQKEAEVLIAKY--TDSLSKANATET-----LAGL--RNMLTLOIM 766
DB 1199 AVPRRRYKQSLKQIRLPLEEIKPERGVVLANATIRNSESSTILOGAGRNNSLFL 1254

```

RESULT 43
US-09-107-532A-5095

Sequence 5095, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 5095:
SEQUENCE CHARACTERISTICS:
LENGTH: 525 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (8) LOCATION 1..525
SEQUENCE DESCRIPTION: SEQ ID NO: 5095:
US-09-107-532A-5095
Query Match 3.1%; Score 129.5; DB 4; Length 525;
Best Local Similarity 21.1%; Pred. No. 0.098;
Matches 93; Conservative 60; Mismatches 186; Indels 101; Gaps 19;
QY 8 QARTVKNRVS-YIDQKATOKTENLTPDESKREG-INAQIVIKITDQGVYTSRSHD 65
DB 108 QAREAGVSTSSNYIDAV-----LNADSLADALGRVQAMTWKXANNMDLMEQCKDK 159
QY 66 YHYNKKVPIYDAISELMLKMDPNYKLBEDYNEKGVIVIKVDGKYVYLLDAHAADN 125
DB 160 KAVEDKKAENDALKE-----LAENQALLESQKGLSKQADLVN-LKTSLAAB 208
QY 126 VRT---KEELNROKQSHREGGTPRNDGAVALARSGRYTTDDGYIFNAAIIEDTGG 182
DB 209 ATREDKADLNKRAE-AEAQKARIEQQLAEQAQQAQAAO-----248
QY 183 AVIVPHGDHYHYIPKNELSASELAAEAFLSGHNSNSRTYRONSNDNTSRTNWVPSV 242
DB 249 -----EKAKEKAREQAEBAQAQTOASTQSSATESSATQSSWTESSSAT 295
QY 243 NPGTNTNTNSNTNSQASQSDISLKLQLYKPLSQSHVSDGIVFPAPQITRTAR 302
DB 296 QSSATEESTTPESSTEESTA-----PESSATEES-----TTAESSATEE 335

QY 303 GAUVPHG---DHYHFIYQSMSELEERIAIIPLAYRS---NHWDSDREOPSPOPTPE 356
DB 336 STVPESASATEESTVPESSSTEESTPTAPTPSTQSDVTGNGTSSSTP-ATPTPTPE 394
QY 357 -RSP-GPQAPNLIKIDNSLSLVQVRKYGEGV-----FEENGISRYF-----AKD 402
DB 395 QPKPVTPAPAPGSGSV-NGAAIVAAYKYGITPVWGGKDPGSGFGFTRVYVMTGRD 453
QY 403 LPSETV--KNLSKLSKQES 420
DB 454 IGGWTVQESAGTKLSVSOA 473
RESULT 44
US-08-826-267-2
Sequence 2, Application US/08826267
Patent No. 5994070
GENERAL INFORMATION:
APPLICANT: Streuli, Michel
TITLE OF INVENTION: No. 5994070el TRIO Molecules and Uses Related Thereto
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: LAHIVE & COCKFIELD
STREET: 28 State Street
CITY: Boston
STATE: Massachusetts
COUNTRY: USA
ZIP: 02109-1875
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/826,267
FILING DATE: 1997
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/014,214
FILING DATE: 27 MARCH (1996)
ATTORNEY/AGENT INFORMATION:
NAME: Amy E. Mandragouras
REGISTRATION NUMBER: 36,207
REFERENCE/DOCKET NUMBER: DFN-010
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)227-7400
TELEFAX: (617)227-5941
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2860 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-826-267-2
Query Match 3.1%; Score 129; DB 2; Length 2860;
Best Local Similarity 19.3%; Pred. No. 1.6; 260; Indels 200; Gaps 36;
Matches 136; Conservative 108; Mismatches
QY 147 TRPDGAVALARSGO---RYTTDDGYIFNADSI---IEDTGAYIYPHGDHYHYIPKNEI 200
DB 140 TPEFDGCLYNNHEWEIIVAFED-YISNATMTLSLELOD-----ILAKKEL 187
QY 201 SASLELAAPALPSGRNL-----SNSRTYRONS-SDNTSRT 235
DB 188 -PDLEGANMIEHSQKLVKAPIEDLDLEGQKLQRIOSSESFPKKNSSGSGNADIQ 246
QY 236 NWPVSNSNGTNTNTNSNSNTNSQASQSDISL-LKQLYKPLSQSHVSDGIVFPD 292
DB 247 NLIPKVS-----TMDRLHSTRQHLQWVHVRKLLDQCFQRLPEQDAEK---MFDWI 297
QY 293 ---PAQITSRTARGAUVPHG---DHYHFIYQSMSELEERIAIIPLAYRS---SNHW 340

```

Db      298  THINKGLFLNYSYETIGSHPEMELQTOHNHFA--NMCMNVVYNINRINSVANRLVESGHY 355
Qy      341  VPDSPREQSPQPTPEPSRGPQAPNLKIDSSSL--VSQVLRKXGEGVFEKGISRYV 398
Db      356  A-SQOIRQASOLEDE---WKAFPAALDERSTLLDSSSIHQAEKMSVDSWCYAC 409
Qy      399  FAKDLPSSETVKNLSEKLSKQESVSHTLTAKKENVAPRDOEFQKXVNLTEAHKALFYXK 458
Db      410  GEVDLPSE-LQDLEDAIHNHGQIYEHITLAYSEVSGDDKSLDKQRPILT-----P 459
Qy      459  GRNSFOALDKLERLNDESTNKEKLVDDLAFAPITPHEKLGKPNQOIEETDEVRIA 518
Db      460  GSSDSLTA-----SANYSKAVHHLVDVIEHVLHQR---HHTTIQHRKVRILH 504
Qy      519  QLADKYTSDGYIPDEH-----DIISDEGDAYVTTHMGSHWIGKOSL-----SD 563
Db      505  QRLQI-----CVFOOEVOQVLDMWENGEAFLSKHTG--VGR-SLRARALQKRHD 553
Qy      564  KEKVAQAQYTKKEGILPSPDADVKANPTGD-----SAAIYNRVKGEKRIPLVRLP 615
Db      554  FEEVQONTYTNADKLL-----EAAEQLAQTECEDPREIYQAAHQLDRIODFVR----- 602
Qy      616  YVHEHTVEKNGNLIIPHKHVH-NIKFAPEDHTYKAPNGYTLDELFTTIKYVVEHPDE 674
Db      603  -----FVEQRKILLDMSVSFHTVYKELM---TW-----LEEL-----Q 632
Qy      675  RPHSNDGCMNASEHVLGKKHSEDPNKNKKADEEVEETPAPEVPOVETEKVEAOKEA 734
Db      633  KELLDVVAESVEAV-----ODLIKRFQOO-----QTTLOVTAVVIVEG 672
Qy      735  EVLLAKVTDSLKANATETLAGLRNNLTLOIMDNNSIMAEKEL 778
Db      673  EDLIQQLRDSAISNKTPHNSSI-NHIEITVLOQLDEAQSOMEL 715

```

```

RESULT 45
US-08-188-228-48
; Sequence 48, Application US/08188228
; Patent No. 5597725
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; STREET: 6300 Sears Tower, 233 S. Wacker Drive
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/188,228
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/049,460
; FILING DATE: 19 APR 1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,643
; FILING DATE: 17 APR 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 5597725and, Greta E.
; REGISTRATION NUMBER: 35,302
; REFERENCE/DOCKET NUMBER: 31340
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448

```

```

; TELEX: 25-3856
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 916 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-188-228-48

```

```

Query Match      3.1%; Score 128.5; DB 1; Length 916;
Best Local Similarity 19.9%; Pred. No. 0.26; Indels 237; Gaps 32;
Matches 132; Conservative 79; Mismatches 214;

```

```

Qy      218  LSNRTYRQNSDNTSRTPNWPVSVPCTT-----NTNNSNNTNSQASQ 263
Db      12  LSLGALRAHNEIDLTRCTKAGFSBDDYTLALQNIIEGKLLQVKESSCVGTGTQYE 71
Qy      264  SNDISLLKQYLKPLSGRHVESDGLVFPDPAQITSRTARGVAVPHGDHYHPIY--SQMS 321
Db      72  TNSMDFL-----VGADGTVF-----ATRELQVSEQVAFVTAMDQTA 110
Qy      322  ELERIAIITPLRYSNHWVDSRPEQ-----PSFOPTPEPSRGPQ----- 362
Db      111  EKWDAVRLVAQTSPPH--SGHKPQKQKVVALDPSPPKDTLLPWFQONANGLRKK 168
Qy      363  -----PAPNLKIDSSSLVSQLVRKVGEGYVFEKGISRYVFAKDLPSE-TVKNLESKLS 416
Db      169  RDWVIFPILNVENSRGPPQOLVR-----IRSKD-----NDIPRYSITVGADQP 215
Qy      417  KQESVS--HTLTAKKENVAPRDOEFYDKAYNLLTEAHKALFYXKGRNSDFOALDKLERLN 475
Db      216  PMEVSISMSGMYVTRPMDREH--ASYHL--RAHAV----- 250
Qy      476  DESTNKEKLVDDLAFAPIT--TPERL-----GKPNQOIEETDEVRIAQIAD 522
Db      251  DMNGNKVENPIDLYIYVDMNDNHEPITNOYNCVDEGSKPGTYV---MTITANDAD 305
Qy      523  KYTTSDDGYI-----PDEHDIISDEGDAYVTPHMGSHWIGKDSLSPKEKVA 569
Db      306  DSTTANGNVRIRIYQTPQPSQNMFTINSETG-IYVVAQ---W-----DREK-- 352
Qy      570  QAYTEKGLPSPD-----ADVKANPTGSAAIYNRVKGEKRI 610
Db      353  QOYT-----VIVQATDMEGNLVGLSNTATAIITVTDVANDSEFTASTFAGEVP----- 402
Qy      611  LVRLPYVHEHTVEKNGNLIIPHKHVHNIKFAFPDHTTYKAPNGYTLDELFTTIKYVE 670
Db      403  -----ENSVETVVALTVMDRD----- 419
Qy      671  HPDRPHSNDGCMNASEHVLGKKHSEDPNKNKKADEEVEETPAPEVPOVETEKVEAQ 730
Db      420  -----QPHS-PNW-NAVYRII-----SGDPSGHFVSVRDPVYN---EGNVTVVKA--VDYE 463
Qy      731  LKEAEVLAKYTDSSLKANATETLAGLRNNLTLOIMDNN--IMAEKLLALLKGSNPS 788
Db      464  LNRFAFLTVWVSNQAPLASGIQMSFOSTAGVTISIMDINEAPYPPSNKILIRLEGEVPG 523
Qy      789  SV 790
Db      524  TV 525

```

```

RESULT 46
US-08-332-643-42
; Sequence 42, Application US/08332643
; Patent No. 5639634
; GENERAL INFORMATION:
; APPLICANT: Suzuki, Shintaro
; TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
; NUMBER OF SEQUENCES: 56
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
; ADDRESS: Bicknell

```

STREET: Two First National Plaza, 20 South Clark
STREET: Street
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/332,643
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/872,643
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: No. 5639634and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 27866/30795
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 346-5750
TELEFAX: (312) 984-9740
TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 916 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-332-643-42

Query Match 3.1%; Score 128.5; DB 1; Length 916;
Best Local Similarity 19.9%; Pred. No. 0.28;
Matches 132; Conservative 79; Mismatches 214; Indels 237; Gaps 32;

218 LNSRTYRQNSDNTSRTNWPVSVPNGTT-----NTNTSNTSNTNSQASQ 263
12 LLSGALRAHNBDLTTRCTCKAGFSEDDYITALISQNLLEGKLLQVKFSSCVGTGCTGYE 71
264 SNDISLKLQLYKLPLSQRHVESDGLVPDPAQITSTRAGVAVPHGDHFIPIY--SQMS 321
72 TNSMDFL-----VGADGTVF-----ATRELQVPSQVAFVTYAMDSSQTA 110
322 ELEBRARIILPLRYSNHWVPDSRPEQ-----PSPOPTPEPSGPGQ----- 362
111 EKMDAVRLVLAQTSSPH--SGHKPKQKGVVALDPSPPKDTLLPWPQHONANGLRKK 168
363 -----PAPULKIDSNSLSVQLVRKVGEGVFEKGISRYVFAKDLPSR-TVQNLESKLS 416
169 RMWVPIPIVNPENSGRPPQQLVR-----INSDKD-----NDIPRYSTGVADGP 215
417 KOESVS-ATLTAKKENAVAPROEFYDKAVNLLTEAHKALFXNKGSRSDFOALDKLERLN 475
216 PHEVTSINMSGRMYTRPMDRREH-ASYHL--RAHAV----- 250
476 DESTNKEKLVLDLAFAPL--THPERL-----GKPSQIEYTEDEVRLAQLAD 522
251 DMNGNVEVNPIDLYIVIDMNDHPEFINQVYVNCVSDEGSKPGTYV-----MTITANDAD 305
523 KYTTSQGY-----FDEHDIISDEGDVAVTTHMGSHHIGKDSLSDKRVAA 569
306 DSTTANGMWRVYIVTQTPQSPSQNMFTINSETGD-IVTAAAG--W-----DREKV- 352
570 QAYTEKGLPSPD-----ADVKANPTGDSAAAIYNNRYKSGEKRIAP 610
353 QGYT-----VIQATDMEGNLNYGLSTATAITIVTDVNNPSPFTSTAGAEVP----- 402
611 LVRLPYMEHTVEVKNGLIIPKQDHYHNKFAAMPDDHTYKAPNGVYTLDELFAITTYVE 670
403 -----ENSVETVAVNLTVMDRD----- 419

671 HPERPHSNDGMASEHYLGKDHSEDPKNKPKADEBEVEETPAPEVPOVETEKVFAO 730
420 -----QPHS-PNW-NAVYRII-----SGDPSGHFVSRTDVTN--EGWTVVKA--VDYE 463
731 LKEAEVYLAKYTDSSIKANATETLAGLRNNLTLOIMDNNS--IMAEKYLALLKGSNPS 788
464 LNRAPMLTVWVNOAPLASGIQMSFOSTAGVITISINDINEARYPSPNHKLIRLEGVPPG 523
789 SV 790
524 TV 525

RESULT 47
US-08-332-638-48
Sequence 48, Application US/08332638
Patent No. 5646250

GENERAL INFORMATION:
APPLICANT: Suzuki, Shintaro
TITLE OF INVENTION: CADHERIN MATERIALS AND METHODS
NUMBER OF SEQUENCES: 62
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray &
STREET: 6300 Sears Tower, 233 S. Wacker Drive
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 60606

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/332,638
FILING DATE: 01-NOV-1994
CLASSIFICATION: 435

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,643
FILING DATE: 17 APR 1992
APPLICATION NUMBER: US/08/049,460

ATTORNEY/AGENT INFORMATION:
NAME: No. 5646250and, Greta E.
REGISTRATION NUMBER: 35,302
REFERENCE/DOCKET NUMBER: 31340
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448

TELEX: 25-3856
INFORMATION FOR SEQ ID NO: 48:
SEQUENCE CHARACTERISTICS:
LENGTH: 916 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-332-638-48

Query Match 3.1%; Score 128.5; DB 1; Length 916;
Best Local Similarity 19.9%; Pred. No. 0.28;
Matches 132; Conservative 79; Mismatches 214; Indels 237; Gaps 32;

218 LNSRTYRQNSDNTSRTNWPVSVPNGTT-----NTNTSNTSNTNSQASQ 263
12 LLSGALRAHNBDLTTRCTCKAGFSEDDYITALISQNLLEGKLLQVKFSSCVGTGCTGYE 71
264 SNDISLKLQLYKLPLSQRHVESDGLVPDPAQITSTRAGVAVPHGDHFIPIY--SQMS 321
72 TNSMDFL-----VGADGTVF-----ATRELQVPSQVAFVTYAMDSSQTA 110
322 ELEBRARIILPLRYSNHWVPDSRPEQ-----PSPOPTPEPSGPGQ----- 362

```

Db 111 EKMDAVRLVLAQTSSPH--SGHKPKGKQKVVALDPSPPKOTLLPWPQHANGLRBRK 168
Qy 363 -----PARNLKIDSSSLVSQVLRKVGEGYVEEKGISRYAKLPSP--TYKULESKLS 416
Db 169 RDWVLPINVPENSSGPPQOQVR-----IRSDKD-----NDIPYRSTTGGAQOP 215
Qy 417 KOESVS-HTLTKAKKNVAPRDOEFYDKANLLTEAHKALFXNKGNSDPQALDKLERLN 475
Db 216 PMEVSINSMGSRMVTTPMDEEH-ASYHL--RAHAV-----GKXSOIEYTEDEVRIAOAD 250
Qy 476 DESTNKEKLVDDLAFLAPI--THEPRL-----GKXSOIEYTEDEVRIAOAD 522
Db 251 DMNGKVENPIDLYIVIDMNDNHEFINQVYNGSVDEGSKGTYY-----MTITANDAD 305
Qy 523 KTTSDGYI-----FDEHDIISDEGDVYTPHMGSHWIGKDISDKEXAA 569
Db 306 DSTTANGVRYRIVTQTPSPSQNMFTINSETGD-IVTVAAQ-----W-----DREKV-- 352
Qy 570 QAYTEKGIPLPSPD-----ADVKANPTGDSAAAIVNRYKGEKRIIP 610
Db 353 QOYT---VIYQATMEGKLYGLSNTATAIITVTDVNDNPEEFTASTFAGEVP----- 402
Qy 611 LVRLPYWEHTVEVNGNLIIPKDHYNHKEFAMEDDHTYKAPNGYTLLEDLPATIKYYVE 670
Db 403 -----ENSVETVVAULTVMDRO----- 419
Qy 671 HPDERPHSNDGKGNASEHVLGKKHSEDPNKFKADEEVPVETPAEPEVPOVETKEVDAQ 730
Db 420 ---OPHS--PMW--NAVYRII-----SGDPSGHFSVRTDPVTN--EGWTVVKA--VVE 463
Qy 721 LKEAEVLAKYVTDSSLKANATETLAGLRNNLFLQIMDNNS--IMAEKLTALLKGSNPS 788
Db 464 LNRALMLTVMSNQAPLASGIOMSFQSTAGVTITISINDINEAPFSPNKLILLEGCVPRG 523
Qy 789 SV 790
Db 524 TV 525

```

RESULT 48
US-08-296-791-6
Sequence 6, Application US/08296791
Patent No. 6245337
GENERAL INFORMATION:
APPLICANT: St. Geme III, Joseph W.
APPLICANT: Falkow, Stanley
TITLE OF INVENTION: Haemophilus Adherence and Penetration
TITLE OF INVENTION: Protein
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert
STREET: 4 Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/296,791
FILING DATE: 25-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Treccartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249

```

/ TELEX: 910 277299
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1848 amino acids
/ TYPE: amino acid
/ TOPOLOGY: unknown
US-08-296-791-6

Query Match 3.1%; Score 128.5; DB 3; Length 1848;
Best Local Similarity 18.6%; Pred. No. 0.86; Indels 325; Gaps 44;
Matches 161; Conservative 109; Mismatches 271;

Qy 1 SYEGLYQART--YKENNRVSVID-----GKQATOKTEN----- 32
Db 632 SNEMLVNGRTSDAKKNVNMHINNERNGNGYFGESETATQNGKLVTFNGKSDQR 691
Qy 33 -LTPDEVSKREGINAEOIVIKITDGYVTSQHDHYHYNGKVPYDAIISELLMKDPNYK 91
Db 692 FLITGCTNLNGDLNVEKGTFLFS--GRPTPHARD-----IAGISSTKKDPHFT 737
Qy 92 -----LKDEDIIVNEKGVYIKVDGKYVYIKDAADNVRTKEINRQKQEHQHRGCG 146
Db 738 ENNEVVEEDDWINNFKATTNVTGNASLY--SGRNVANITSNTTASNNAGVHIQYKTD 795
Qy 147 TPRNDGAVALARSQGRYTTDDGYI-----FNASDI----- 176
Db 796 T-----VCYRSD--YT--GYTCHNSNLSEKALNSFNTNLKGNVNLTENASFTLGK 843
Qy 177 -----IEDTGDVAYI-VPHGHHYIIPKNELASASELAAEAFLSGRGNLSRSTRYRONS 229
Db 844 ANLFGTIOISIGTSQVNLKNSHMLTGNSVNLQNLTGCHILHNAQDANKVTYI----- 898
Qy 230 DNTSRTVNVPSVNSPQT-----TNTNTSNSTNTNSQASQNSNDIDSLKQLYLPLSQR 282
Db 899 -NTLTVN--SLSSNGSYFYWDFTN--NKSQVNVNKSATGN-----FTLQVADK 943
Qy 283 HVE---SDGLVDFDPAQITSR---TARGVAVPHG-----DHYHFIYQSOWSE 322
Db 944 TGEPNHNEHLTFDASNAITRNMLEVTLANGSVDRGAMKYKLRNVNGRYDLVN-----PE 996
Qy 323 LEERIARIIPLYRSNHWVP--DSRPEQSPQPTPE-----DSPGPQAPANIKIDSNSSLV 376
Db 997 VEKENQTV---DTTNITTPMDIQADAPSAQSNNEEJARVETPVPPAP----- 1041
Qy 377 SOLVRKNGEGVYFPEKGSRYVFAKDLPSFVKNVLESLSKQESVSHLTAKENVAPRD 436
Db 1042 -----ATESAIASEQPETR-----PAETA-----QPMEEETNTANSTETAPK- 1078
Qy 437 QEFYDKAVNLLTEAHKALFXNKGNSDPQALDKLERLNDESTYKEXLVDDLAFLAPI 496
Db 1079 -----SDTATQTE----- 1086
Qy 497 HPERLGRNSQIEYTEDEVRIAOADKYTSDGYIFDEHDIISDEGDA-VYTPHMGSHW 555
Db 1087 -----NPNSSEVPSETTEKVAENPQ-----ENETVAKKQEQETETEPONGE--- 1128
Qy 556 IGRDLSLDEKVAQAQATYK-----KGIPLPSPDADVANKPTGDSAAAIVNRYKGEKRIPL 611
Db 1129 VAKK---DQPTVEANTQNEATQSEKTEFTQTAETSEPT--ESVTVSENG----- 1175
Qy 612 VRLPYWEHTVEVNGNLIIPKDHYNHKEFAMEDDHTYKAPNGYTLLEDLPATIKYYVEH 671
Db 1176 -----PEKTVSOSTEDKVVVEKEKAKVE-----TEETQKAPQVTSKE-----PKQAEPA 1221
Qy 672 PDERPHSNDGKGNASEHVLGKKH-----SEDPNKFKADEE--PVEETPAEPEVPOV 722
Db 1222 PEVVPYDT---NAEEQAALQOQTPPTVAAAEITSPNSKPAEFTQOQSEKTNABEVPVTV 1277
Qy 723 -----ETEKVBAQJKEAEV 736
Db 1278 SENTATQPTETETETAKVEKE-KTQEV 1302

```



```

Db 237 LYKSGLSRQAQNSNAVENDLP--VGPG-----AAGPSNVPAFLTKLWT- 278
QY 278 PLUSQHVESDGLVFPDPAQITSTARGAVAPHGDHYFIPIYQMSLEERIAIIPLRYS 337
Db 279 -----LVSDP---DTDALICWSPSGNSFHFVFDQGFQAK-----EVLKPYFQH 317
QY 338 NHMVDSRPEQSPQPTPEPSPGPQAPAPYLKIDNSSLVSQL-----VRKVG-----GYVF 389
Db 318 NNM-----ASFVRQLMYGFRRKVHIEQGLVK 345
QY 390 BEKGISRYVFAKL--PSETVKNLESKLSKQESVSHLTAKKENVAPRDQEFYDQAVNLL 447
Db 346 PERDTEFQHPCEFLRQEQLEENIKRKVT---SVS---TLKSEDIKIRO---DSVTKL 395
QY 448 TEAH-----KALFXNKGNSDFQALDKLERLNDESTNKEXLVDDLLAFLAPITH 497
Db 396 TDVOIMKQKQCBMSKLLAMKHEN--EALWREVASLRQKHAQQKVVNKLIOPLISLVQ 452
QY 498 PERLGKPNQSOLEYTEDEVRIAQLADKYTTSQGYIFDEHDIISDEGDAYVTP-----HMG 551
Db 453 SNRIIGVKKRKLPL-----MUNDGSAHSMFKYSRQFSLE 486
QY 552 HSHWIGKDSLSDKEKVAQAQYTKKGIPLPSPDADVKANPTGDSAAAIYNNRYKGEKRIPL 611
Db 487 HVHSGGYPASAPSPAYSSSLY-----APDAVASSGPI----- 518
QY 612 VRLPYMVEHTVEVKNQNLIIPIKDHVHNKIFAMFDDHTYKAPNGYTLLEDLPATIKYVEH 671
Db 519 -----ISDITELAP----- 527
QY 672 PDERPHSNDGMCNASEHVLGKKHSEDPNKFNKADBEPEETPAPEVEPOVE----- 723
Db 528 --ASPMASPG-GSIDERPL-----SSSPLYRVK-----EEPPSPQSPRVVEASPGRPS 573
QY 724 -----TEKVEAQLKEAEVLLAKVTDSSLKANATET 753
Db 574 SVDTLSPALIDSILRESEPAASVTALTDAARGHTDT 611

```

Search completed: November 14, 2003, 10:38:25
 Job time : 52 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2003, 10:37:50 ; Search time 332 Seconds
(without alignments)
437.703 Million cell updates/sec

Title: US-09-765-271-56

Perfect score: 4165
Sequence: 1 SYELGLYQARTVKNRVS.....KLALLKGSNPSVSXKIN 796

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 666188 seqs, 182559486 residues

Total number of hits satisfying chosen parameters: 666188

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : Published Applications AA.*

```
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4163	100.0	796	US-09-765-272-56	Sequence 56, Appli
2	3218	77.3	840	US-09-884-465A-7	Sequence 7, Appli
3	2795	67.1	826	US-09-769-787-194	Sequence 194, Appl
4	2750.5	66.0	838	US-09-884-465A-8	Sequence 8, Appli
5	2649.5	63.6	763	US-09-765-272-66	Sequence 66, Appli
6	1899	45.6	1126	US-09-884-465A-383	Sequence 383, Appl
7	1891.5	45.4	1365	US-09-884-465A-382	Sequence 382, Appl
8	1795.5	43.1	1139	US-09-884-465A-380	Sequence 380, Appl
9	1795.5	43.1	1378	US-09-884-465A-378	Sequence 378, Appl
10	1665	40.0	1238	US-09-884-465A-381	Sequence 381, Appl
11	1656	39.8	989	US-09-884-465A-376	Sequence 376, Appl
12	1649	39.6	999	US-09-884-465A-377	Sequence 377, Appl
13	1247	29.9	1039	US-09-884-465A-6	Sequence 6, Appli
14	1203.5	28.9	484	US-09-769-787-38	Sequence 38, Appli
15	1203.5	28.9	484	US-09-769-744A-24	Sequence 24, Appli

16	1187.5	28.5	447	9	US-09-765-272-182	Sequence 182, App
17	991.5	23.8	913	11	US-09-884-465A-384	Sequence 384, App
18	991.5	23.8	1152	11	US-09-884-465A-379	Sequence 379, App
19	966.5	23.2	840	11	US-09-884-465A-10	Sequence 10, Appli
20	929	22.3	822	12	US-09-769-736-18	Sequence 18, Appli
21	927	22.3	793	11	US-09-252-088-15	Sequence 15, Appli
22	841	20.2	381	12	US-09-769-736-24	Sequence 24, Appli
23	723.5	17.4	906	11	US-09-884-465A-369	Sequence 369, App
24	714.5	17.2	906	11	US-09-884-465A-371	Sequence 371, App
25	710.5	17.1	906	11	US-09-884-465A-373	Sequence 373, App
26	673.5	16.2	900	11	US-09-884-465A-333	Sequence 333, App
27	660	15.8	715	11	US-09-252-088-16	Sequence 16, Appli
28	645	15.5	906	11	US-09-884-465A-370	Sequence 370, App
29	645	15.5	906	11	US-09-884-465A-372	Sequence 372, App
30	643	15.4	272	11	US-09-884-465A-258	Sequence 258, App
31	643	15.4	895	11	US-09-884-465A-344	Sequence 344, App
32	643	15.4	901	11	US-09-884-465A-343	Sequence 343, App
33	641	15.4	272	11	US-09-884-465A-300	Sequence 300, App
34	641	15.4	272	11	US-09-884-465A-301	Sequence 301, App
35	636	15.3	894	11	US-09-884-465A-336	Sequence 336, App
36	636	15.3	900	11	US-09-884-465A-335	Sequence 335, App
37	636	15.3	906	11	US-09-884-465A-332	Sequence 332, App
38	634	15.2	272	11	US-09-884-465A-293	Sequence 293, App
39	634	15.2	272	11	US-09-884-465A-294	Sequence 294, App
40	634	15.2	272	11	US-09-884-465A-295	Sequence 295, App
41	627	15.1	894	11	US-09-884-465A-340	Sequence 340, App
42	627	15.1	894	11	US-09-884-465A-342	Sequence 342, App
43	627	15.1	900	11	US-09-884-465A-339	Sequence 339, App
44	627	15.1	900	11	US-09-884-465A-341	Sequence 341, App
45	626	15.0	272	11	US-09-884-465A-302	Sequence 302, App
46	625	15.0	272	11	US-09-884-465A-296	Sequence 296, App
47	625	15.0	272	11	US-09-884-465A-298	Sequence 298, App
48	625	15.0	272	11	US-09-884-465A-299	Sequence 299, App
49	625	15.0	895	11	US-09-884-465A-346	Sequence 346, App
50	625	15.0	901	11	US-09-884-465A-345	Sequence 345, App
51	616	14.8	272	11	US-09-884-465A-303	Sequence 303, App
52	616	14.7	268	11	US-09-884-465A-304	Sequence 304, App
53	598	14.4	259	11	US-09-884-465A-374	Sequence 374, App
54	597	14.3	888	11	US-09-884-465A-338	Sequence 338, App
55	597	14.3	894	11	US-09-884-465A-337	Sequence 337, App
56	597	14.3	900	11	US-09-884-465A-304	Sequence 304, App
57	595	14.3	266	11	US-09-884-465A-304	Sequence 304, App
58	595	14.3	889	11	US-09-884-465A-348	Sequence 348, App
59	595	14.3	895	11	US-09-884-465A-347	Sequence 347, App
60	592	14.2	268	11	US-09-884-465A-375	Sequence 375, App
61	512.5	12.3	485	12	US-09-769-736-72	Sequence 72, Appli
62	444	10.7	94	11	US-09-884-465A-11	Sequence 11, Appli
63	302	7.3	86	11	US-09-884-465A-22	Sequence 22, Appli
64	213	5.1	42	11	US-09-884-465A-12	Sequence 12, Appli
65	176	4.2	41	11	US-09-884-465A-23	Sequence 23, Appli
66	170	4.1	69	11	US-09-884-465A-24	Sequence 24, Appli
67	170	4.1	2478	9	US-09-815-242-5816	Sequence 5816, App
68	170	4.1	2478	9	US-09-815-242-12967	Sequence 12967, A
69	169	4.1	886	8	US-08-781-986A-5235	Sequence 5235, App
70	169	4.1	909	12	US-10-172-502-6	Sequence 6, Appli
71	168.5	4.0	1043	10	US-09-946-805-4	Sequence 4, Appli
72	159	3.8	870	9	US-09-815-242-5493	Sequence 5493, App
73	159	3.8	870	9	US-09-815-242-1637	Sequence 12637, A
74	158	3.8	1118	15	US-10-153-668-104	Sequence 104, App
75	156	3.7	519	11	US-09-934-445-164	Sequence 164, App
76	151.5	3.6	2086	9	US-09-815-242-5639	Sequence 5639, App
77	151.5	3.6	5795	9	US-09-815-242-12610	Sequence 12610, A
78	151	3.6	1610	15	US-10-153-533-9	Sequence 9, Appli
79	147	3.5	812	15	US-10-100-957A-6	Sequence 6, Appli
80	145.5	3.5	2668	9	US-09-815-242-5635	Sequence 5635, App
81	145.5	3.5	2668	9	US-09-815-242-12389	Sequence 12389, A
82	144	3.5	1040	9	US-09-765-272-1118	Sequence 118, App
83	143	3.4	633	11	US-09-884-465A-350	Sequence 350, App
84	143	3.4	633	11	US-09-884-465A-354	Sequence 354, App
85	143	3.4	633	11	US-09-884-465A-361	Sequence 361, App
86	143	3.4	633	11	US-09-884-465A-362	Sequence 362, App
87	143	3.4	2025	9	US-09-815-242-5703	Sequence 5703, App
88	143	3.4	2659	12	US-10-311-879-28	Sequence 28, Appli

89 143 3.4 3158 9 US-09-815-242-12611 Sequence 12611, A
90 142.5 3.4 6281 9 US-09-815-242-12996 Sequence 12996, A
91 141.5 3.4 2314 11 US-09-983-000A-6 GENERAL INFORMATION
92 141.5 3.4 2353 11 US-09-983-000A-4 Sequence 4, Appli
93 141.5 3.4 2437 9 US-09-815-242-5834 Sequence 5834, Ap
94 141 3.4 2431 10 US-09-901-106-2 Sequence 2, Appli
95 140 3.4 1111 9 US-09-815-242-12955 Sequence 12955, A
96 140 3.4 1637 12 US-10-172-502-14 Sequence 14, Appl
97 139.5 3.3 1576 15 US-10-037-182-16 Sequence 16, Appl
98 139.5 3.3 1609 10 US-09-938-275-11 Sequence 11, Appl
99 139.5 3.3 1609 15 US-10-037-182-14 Sequence 14, Appl
100 139.5 3.3 1609 15 US-10-299-058-12 Sequence 12, Appl

ALIGNMENTS

RESULT 1

US-09-765-272-56

Sequence 56, Application US/09765272

Patent No. US20020061545A1

GENERAL INFORMATION:

APPLICANT: Choi et. al.

TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines

NUMBER OF SEQUENCES: 452

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.

STREET: 9410 Key West Avenue

CITY: Rockville

STATE: Maryland

COUNTRY: USA

ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage

COMPUTER: HP Vectra 486/33

OPERATING SYSTEM: MSDOS version 6.2

SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/765,272

FILING DATE: 22-Jan-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/961,083

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Brookes, A. Anders

REGISTRATION NUMBER: 36,373

REFERENCE/DOCKET NUMBER: PB340P2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (301) 309-8504

TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 56:

SEQUENCE CHARACTERISTICS:

LENGTH: 796 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 56:

US-09-765-272-56

Query Match

Best Local Similarity 100.0%; Score 4163; DB 9; Length 796;

Matches 796; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SYELGLYQARTVKNRRVSYIDGKQATQKTENLTPDEVSKREGINAEGIVIKITDQGYVT 60
DB 1 SYELGLYQARTVKNRRVSYIDGKQATQKTENLTPDEVSKREGINAEGIVIKITDQGYVT 60
QY 61 SHGDHYHYNGKVPYDAIISELLMKDPYKLDKEDIYNEVGKGYIVIKYDGKYYVYLKDA 120
DB 61 SHGDHYHYNGKVPYDAIISELLMKDPYKLDKEDIYNEVGKGYIVIKYDGKYYVYLKDA 120

QY 121 AHADNRTKEINRQKQESHQREHGTGPRNDGAVALARSOGRYTTDDGYIFNADIEDT 180
DB 121 AHADNRTKEINRQKQESHQREHGTGPRNDGAVALARSOGRYTTDDGYIFNADIEDT 180
QY 181 GDVIVPHGDHYHYIIPKQELASASLAAEAFLSRGULNSRTYRRONSNTSTNNWPS 240
DB 181 GDVIVPHGDHYHYIIPKQELASASLAAEAFLSRGULNSRTYRRONSNTSTNNWPS 240
QY 241 VSNNGTTNTNTSNNSNTNSQASQNSNDIDSLKQLYKPLPSORHVESGLVDPQOITSRT 300
DB 241 VSNNGTTNTNTSNNSNTNSQASQNSNDIDSLKQLYKPLPSORHVESGLVDPQOITSRT 300
QY 301 ARGVAVPHGDHYHYIIPYQMSLEBETARIILPRYSRSHWVPSRPEQSPQTPPEPSG 360
DB 301 ARGVAVPHGDHYHYIIPYQMSLEBETARIILPRYSRSHWVPSRPEQSPQTPPEPSG 360
QY 361 POPAPNLKIDNSSLVQLYRKVGEVFEKGISRYVFAQDLPSFTYKULESKLKSQES 420
DB 361 POPAPNLKIDNSSLVQLYRKVGEVFEKGISRYVFAQDLPSFTYKULESKLKSQES 420
QY 421 VSHLTAKKENVAPRDEPFYDKAYNLLTEAHKALFXNKGKNSDPQALDKLERLNDESTN 480
DB 421 VSHLTAKKENVAPRDEPFYDKAYNLLTEAHKALFXNKGKNSDPQALDKLERLNDESTN 480
QY 481 KEKLVDDLAFLEITPBERLQKPNQIETYEDEVRIAQLADKYTSDGYTFDEHDIISD 540
DB 481 KEKLVDDLAFLEITPBERLQKPNQIETYEDEVRIAQLADKYTSDGYTFDEHDIISD 540
QY 541 EGDAYVTPHMGSHWIGKDSLSDEKVAQAQYTEKGIILPSPPADVYKANTGSAAIY 600
DB 541 EGDAYVTPHMGSHWIGKDSLSDEKVAQAQYTEKGIILPSPPADVYKANTGSAAIY 600
QY 601 NRVGKERIPLVRLPYWVEHTEVYKNGLLIIPKHQHNIFAFMFDDHTYKAPNGYTLED 660
DB 601 NRVGKERIPLVRLPYWVEHTEVYKNGLLIIPKHQHNIFAFMFDDHTYKAPNGYTLED 660
QY 661 LFATIKYVEHPDRPHSNDGWSASEHVLGKQHSDEPNKFPADPEVETPAEPEVP 720
DB 661 LFATIKYVEHPDRPHSNDGWSASEHVLGKQHSDEPNKFPADPEVETPAEPEVP 720
QY 721 QVETEKYEAQLKEAEVLLAKYTDSLLKANATETLAGLRNNLTLOIMNNSIMAAEKLLA 780
DB 721 QVETEKYEAQLKEAEVLLAKYTDSLLKANATETLAGLRNNLTLOIMNNSIMAAEKLLA 780
QY 781 LKGSNPSVSKEKIN 796
DB 781 LKGSNPSVSKEKIN 796

RESULT 2

US-09-884-465A-7

Sequence 7, Application US/09884465A

Publication No. US20030077293A1

GENERAL INFORMATION:

APPLICANT: Shire Biochem, Inc.

APPLICANT: Hamel, Joe

APPLICANT: Brodeur, Bernard

APPLICANT: Martin, Denis

APPLICANT: Charland, Nathalie

APPLICANT: Ouellet, Catherine

TITLE OF INVENTION: Streptococcus Antigens

FILE REFERENCE: 055190-0044

CURRENT APPLICATION NUMBER: US/09/884,465A

CURRENT FILING DATE: 2001-06-20

PRIOR APPLICATION NUMBER: 60/212,683

PRIOR FILING DATE: 2000-06-20

NUMBER OF SEQ ID NOS: 384

SOFTWARE: Patent in version 3.1

SEQ ID NO 7

LENGTH: 840

TYPE: PRT

ORGANISM: Streptococcus pneumoniae

US-09-884-465A-7


```

RESULT 4
US-09-884-465A-8
; Sequence 8, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biotech, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 8
; LENGTH: 838
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-8

Query Match      66.0%; Score 2750.5; DB 11; Length 838;
Best Local Similarity 64.5%; Pred. No. 1.1e-189;
Matches 538; Conservative 90; Mismatches 145; Indels 61; Gaps 8;

QY 1 SYELGLYOARTV-KENNRVSYIDGKQATOKTENLTPDEVSKREGINAEQIVIKITDGYV 59
DB 21 SYELGRHQAQGVKESNRVSYIDGQAGQAKENLTPDEVSKREGINAEQIVIKITDGYV 80
QY 60 TSHGDHYHYNGKVPYDAIISEELMKDPNYKLKEDIIVNEVGKGYIVIKDGYVYVVKD 119
DB 81 TSHGDHYHYNGKVPYDAIISEELMKDPNYQLKDSIDVNEIKGKGYIVIKDGYVYVVKD 140
QY 120 AAHADNVRTKEEINRQKQESHQREGGTFRNDGVALALARSQGRYTTDGYIFNADIIED 179
DB 141 AAHADNVRTKEEIKRQKQESHSHNHS---RADNVAALARAQGRYTTDGYIFNADIIED 197
QY 180 TGDATVPHGCHYHYIIPKNELSASELAAEAFLSGRNLNSRTYRQNSDNTSRTMWP 239
DB 198 TGDATVPHGCHYHYIIPKNELSASELAAEAFLSGRNLNSRTYRQNSDNTSRTMWP 246
QY 240 SVSNPGTNTNTSNNSNTNSQASQNDIDSLKQLYKLPSQRHVESDGLVFPDPAQITSR 299
DB 247 NPQVPRLSNNHLTVTPYHQ-NQGENISLRLRELYAKPLSERHVESDGLIFDPAQITSR 305
QY 300 TARGVAAPHGCHYHYIIPYSQMSLEERTARIIPLARNSHWVPPDRPQSPQTPPEPS- 358
DB 306 TARGVAAPHGCHYHYIIPYEQMSLEERTARIIPLARNSHWVPPDRPQSPQTPPEPS- 365
QY 359 -----PGQOPANLKIDNSSLSLVKRVGEGYFEFGKISRYYAFADLSEETKYNLES 413
DB 366 SLOGAPNPQAPSPNID--EKLKVAAYKRVGDGVIFFENGYSRYIPAKDLSAETIAGIDS 423
QY 414 KLSQESVSHTLTAKENVAPRDOEFYDKAVNLLTEAKALFXNKGNSDFOALDKLIER 473
DB 424 KLSQESVSHTLTAKENVAPRDOEFYDKAVNLLTEAKALFXNKGNSDFOALDKLIER 483
QY 474 LNDESTKEKLVLDLALAPITHEPERLGKPNSEQEYEDERVIAQLADKTTSGYIFD 533
DB 484 LKDVSSDVKLVLDLALAPIRHBERLGKPNQAQITTYDDEIOVAKLAGKYTEDEGYIFD 543
QY 534 EHDIIISDEGDYVTFPHMGSHWIGKDSLSDEKVAQAQYTERKGLPSPADAVKANPTG 593
DB 544 PRDITSDGDAIVTFPHMTHSHWIKDSSLSEARAAQAQYTERKGLPSPADAVKANPTG 603
QY 594 DSAAIYNRVKGKRIPLVRLPYNVEHTVEYVNGNLLIPKQDHYNIKFAFDDHTYKAP 653
DB 604 KGAETIYNRVKAQKVPIDRMPYNLQYTVVEYVNGSLIIPHYDHYNIKEFWFDEGLYAP 663

```

```

QY 654 NGYTLEDLFAITKYYVEHPDERPHSNDQGNASEHVLCK-----KHSE----- 697
DB 664 KGYSLIEDLATVKKYVEHPDERPHSNDQGNASEHVLCK-----KHSE----- 723
QY 698 -----DPKNFKADEPVEETPAEPEVQVETEKVEAQLKEAEV 736
DB 724 HPSDEKENHAGLNPSADNLKFPSTDEETEEBAEDTDEAEIIPQVENSVINAKIADAEA 783
QY 737 LIAVTDSLSKAAETETLAGLRNNLTQIMNNSIMAEKULALLKGSNSSV 790
DB 784 LLEKVTDPISIRQNMETLTGKSSLLGTQDNNTISAEVSLALLKESQAP 837

RESULT 5
US-09-765-272-66
; Sequence 66, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 66:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 763 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 66:
US-09-765-272-66

Query Match      63.6%; Score 2649.5; DB 9; Length 763;
Best Local Similarity 66.8%; Pred. No. 1.8e-182;
Matches 519; Conservative 77; Mismatches 132; Indels 49; Gaps 8;

QY 1 SYELGLYOARTV-KENNRVSYIDGKQATOKTENLTPDEVSKREGINAEQIVIKITDGYV 59
DB 2 SYELGRHQAQGVKESNRVSYIDGQAGQAKENLTPDEVSKREGINAEQIVIKITDGYV 61
QY 60 TSHGDHYHYNGKVPYDAIISEELMKDPNYKLKEDIIVNEVGKGYIVIKDGYVYVVKD 119
DB 62 TSHGDHYHYNGKVPYDAIISEELMKDPNYQLKDSIDVNEIKGKGYIVIKDGYVYVVKD 121
QY 120 AAHADNVRTKEEINRQKQESHQREGGTFRNDGVALALARSQGRYTTDGYIFNADIIED 179
DB 122 AAHADNVRTKEEIKRQKQESHSHNHS---RADNVAALARAQGRYTTDGYIFNADIIED 178

```

```

0Y 180 GDAIYVPHGDHYHYI PKHELSSELEAAAEAFJLSGNGJNSNRTYRONSNDTNSRTMWP 229
Db 179 TGDATVPHGDHYHYI PKHELSSELEAAAEAFJNSGNGJNSNRTYRONSNDTNSRTMWP 227
0Y 240 SVSNBGTNTNTSNNSNTNSQASQSDNDIDSLKQLYKLPLSQRHVESDGLVPFPAQITSR 239
Db 228 NPAQRLSENHNLVTJWPTVHQ--NQGENISLILRELVAKPLSEKHVESDGLIFDPAQITSR 286
0Y 300 TARGVAVPHGDHYHYI PYQMSLELRIARIITPLRFSNMWVDSRPEQSPDPTPEPSP 359
Db 287 TARGVAVPHGNNHYHYI PYQMSLELRIARIITPLRFSNMWVDSRPEQSPDPTPEPSP 346
0Y 360 GPQAPNLK-IDSN--SSLVSOLYKVGEGYFEEKGTSRYVPADLPSETYKNLESKL 415
Db 347 SPQRPAPNPQAPSNPRLDELKLYEAKRVKQDGYFEEENGSRYYIPAKDLSAETGAAGIDSKL 406
0Y 416 SKQESVSHTLVAKKENVAAPRQDEFYDKAYNLLTEAHKALFXNKGNSDPQALDKLERLN 475
Db 407 AKQESLSHKLGAQKTDLPSSDRDEFYKAYDILLARIHQDLDLNGKRGQVDFEALDNLLERLK 466
0Y 416 DESTKEKXIVDDLFLAFLITTPERLCKNSQIEYEDERINQOLADKYTTSQYIFDEH 555
Db 467 DVXSKRVKXVXDILFLAPIRHPERLCKPAAQITTYDDEIQVAKLKGKTYTDEGYIFDPR 526
0Y 536 DIISDEGDYVYVPHMGSHMWIGKDSLSDEKVAQVAYTEKGLPSPRDADVANKPTGDS 595
Db 527 DITSDEGDYVYVPHMTHSMWIKKDSLSSEAPRAQAQVAAKEKGLTPSTHQBDSGNTBAKG 566
0Y 596 AAATYNRVKGKRIPLVRLPYVEHTVEYKNGNLIIPHKOYHNIKFAWEDHTTYAPNG 6555
Db 587 ABAIYNRVKAARKVPLDRMPYVNLQYTVVEYVNGSLIIPHYDHYNIKEWPFDEGLYEAPKG 646
0Y 656 YLTELPLATIKYVVEHDERPHSNDGMASEHVLGKQHSBDPNNKFKADEE-----P. 709
Db 647 YLTELPLATIKYVVEHDERPHSDNGFNASDHDVQNKQGAQTNQTEKPESEKPOTEKP 706
0Y 710 VEET-----PAPEVPEQVETEKEVAQLKEAEVTLAKVTD 743
Db 707 BEETREKPKQSEKPESPKPTBEPEBESPESEBEPQVETEKVEKLEAEADLLKTIQD 763

RESULT 6
US-09-884-465A-383
; Sequence 383, Application US/09884465A
; Publication No. US2003007293A1
GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
CURRENT FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 383
LENGTH: 1126
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
NAME/KEY: MISC_FEATURE
LOCATION: (1)..(1)
OTHER INFORMATION: Xaa = Methionine or nothing
NAME/KEY: MISC_FEATURE
LOCATION: (557)..(557)
OTHER INFORMATION: Xaa = Glycine or nothing

```

```

; NAME/KEY: MISC FEATURE
; LOCATION: (558)..(558)
; OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-383

Query Match      45.6%, Score 1899; DB 11; Length 1126;
Best Local Similarity 64.9%; Pred. No. 4,7e+128;
Matches 366; Conservative 63; Mismatches 101; Indels 34; Gaps 3

QY 266 DISLKLQLYLPLSQRHVESDGLVDPDAQITSRTARGVAAPHGCHYHFIPLYSONSELE 325
    |||||
DB 2 DIDLKQLLYKLPLSQRHVESDGLIFDPAQITSTRARGVAAPHGHHYFIPYEOMSELEK 61
    |||||

QY 326 RIATIIPLRYSNNMVPDRPREQSPOTTPRPSPPOCAPNIK-IDSN---SLYSOLVR 381
    |||||
DB 62 RIAIIPLRYSNMWPDSPREESPPQTPEPSPPOCAPNFPQAPSPNIDELKYEA VR 121
    |||||

QY 382 KYGEYGFEEKGIRRYFAKDLPSETVNKLSKSIOEVSHTLLPAKKENVA PRROEFYD 441
    |||||
DB 122 KYGGGYFFENGVSRRYPRAKULSABTAAGDISKLAKOESLSHKLGAKKTDLPS SREFYN 181
    |||||

QY 442 KAYNLITFAHKALFXNKGRNSDFQALDKLRLNDESTNKEKLVDDLAFIAPITHPERL 501
    |||||
DB 182 KAYDLLARIHQDLLDNKGRQVDPEALDWLERLKNOVSSDYKLVDDILAFIARHPERL 241
    |||||

QY 502 GKPNISOLEYTDEVRIFAQADKTYTSOGYITFDEHDIIISDEGDAYVT PMGHSHWI GKDSL 561
    |||||
DB 242 GKPAQAITYTDDELIQVAKLACKYTTBEGYIDPRDITSDEGAAYVT PHMTSHHWIKDSL 301
    |||||

QY 562 SDKEKVAAQAATRKKGILPRSPDADVKANPTGDSAAAIYNRVKGEKRIPVRLPYMV EHT 621
    |||||
DB 302 SEARRAAAQAATAKRGKLPBSTDHODSGNTAKGAEALYNRKAAKVP LDMMYNLQTT 361
    |||||

QY 622 VEVANGNLIIPKHCHYHNLIKAWPEDDHYKAPNGYTLEDLPATIKYYVEHPDERPHSNG 681
    |||||
DB 362 VEVGNGSLIIPHVDYHNHIKEFWEDGELYEARPKGYTLEDLATVKKYYEHPRERHSNG 421
    |||||

QY 682 WGNASERYVLGKKHSEDPNKFKAD-----EEPYE 711
    |||||
DB 422 FGNASDHVRKNKOADPTNQTEKSESSEKPOTEKEPEETPRBEKPOSEKPE SPKPTPEERE 481
    |||||

QY 712 ETAPAEPPQVETEKEVEAQLEAVEVLAKVYDSSLKANATELTIGLNUNTLQIMDN NSI 771
    |||||
DB 482 ESPESEEPQVETEKEVEKULEAEADLGKIDPIITSKNAKETLGLKNUNLFGTODNN TI 541
    |||||

QY 772 MAEAEKLIALLKGSNPSSVSKEKI 795
    |||||
DB 542 MAEAEKLIALLKESKXDXDTBEQI 565
    |||||

RESULT 7
US-09-884-465A-382
; Sequence 382, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US-09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,663
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 382
; LENGTH: 1365
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

[illegible]

```

FEATURE:
OTHER INFORMATION: Unknown Organism
NAME/KEY: MISC FEATURE
LOCATION: (1)..(1)
OTHER INFORMATION: Xaa = Methionine or nothing
NAME/KEY: MISC FEATURE
LOCATION: (557)..(557)
OTHER INFORMATION: Xaa = Glycine or nothing
NAME/KEY: MISC FEATURE
LOCATION: (558)..(558)
OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-382

```

```

Query Match      45.4%; Score 1891.5; DB 11; Length 1365;
Best Local Similarity 65.0%; Pred. No. 2.2e-127;
Matches 367; Conservative 60; Mismatches 99; Indels 39; Gaps 4;

```

```

QY 266 DIDLKQLYKLPISORHVESDGLVDPDAQITSTRAGVAVPHGDHYHFIPIYSQMSLEEE 325
DB 2 DIDLKQLYKLPISORHVESDGLVDPDAQITSTRAGVAVPHGDHYHFIPIYSQMSLEEK 61
QY 326 RIARIIPLRYSNMWVPSRPEQSPQTPPEPSPOPAFNPIDKLVKEAVR 121
DB 62 RIARIIPLRYSNMWVPSRPEQSPQTPPEPSPOPAFNPIDKLVKEAVR 121
QY 382 KYGEGYVEEKGISRYVPAKDLPSSETVKNLEKSKOSVSHTLTAKKENVAPROEFYD 441
DB 122 KYGDDYVEEENGVSRYIIPAKDLISASTAGIDSKLAKQESLSHKLGAKTDLPSDREFYN 181
QY 442 KAYNLLTEAHKALFXNKGNSDFOALDKLERLNDSTNKEKLVDDLAFIAPITHPERL 501
DB 182 KAYDILARIHODLLNKGROVDFEALDNLRLKVOSSDKVCLVDDILAFIAPIRHPRL 241
QY 502 GKPNQIETEDVEVIAQLADKYTSDGYIFPEHDIISDEGDAYVTPHMGSHWTKGD 561
DB 242 GKPNQIETEDVEVIAQLADKYTSDGYIFPEHDIISDEGDAYVTPHMGSHWTKGD 301
QY 562 SDKEVNAQAQYKKEGILPPSPDADYKANPTGDSAAIYNRYKGEKRIPLVLPYVME 621
DB 302 SEAEKRAAAQYAKKEGILPPSPDADYKANPTGDSAAIYNRYKGEKRIPLVLPYVME 361
QY 622 VEKNGNLIIPKDHYNHIFKAMPDDHTYKAPNGYTLDELFTIKYVYEHDPDERPHSNDG 681
DB 362 VEKNGNLIIPKDHYNHIFKAMPDDHTYKAPNGYTLDELFTIKYVYEHDPDERPHSNDG 421
QY 682 WGNASEHYLGKXHSDEPNKFKAD-----EEVEE 711
DB 422 FGNASDHVQRNKGADNTQTEKPESEKPOTEKPEEETPREKPOSEKPESEPKPTEEBEE 481
QY 712 ETPAPEPEVOTEKEVEAQLKEAYLAKVYDSSLKANATEETLAGLRNNLTLOIWDNNSI 771
DB 482 ETPAPEPEVOTEKEVEAQLKEAYLAKVYDSSLKANATEETLAGLRNNLTLOIWDNNSI 541
QY 772 MAEAEKLLALKGS-----NPSVS 791
DB 542 MAEAEKLLALKESKXNMQPSQLS 566

```

```

RESULT 8
US-09-884-465A-380
; Sequence 380, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigenes
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20

```

```

PRIOR APPLICATION NUMBER: 60/212,683
PRIOR FILING DATE: 2000-06-20
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 380
LENGTH: 1139
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Unknown Organism
NAME/KEY: MISC FEATURE
LOCATION: (1)..(1)
OTHER INFORMATION: Xaa = Methionine or nothing
NAME/KEY: MISC FEATURE
LOCATION: (570)..(570)
OTHER INFORMATION: Xaa = Glycine or nothing
NAME/KEY: MISC FEATURE
LOCATION: (571)..(571)
OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-380

```

```

Query Match      43.1%; Score 1795.5; DB 11; Length 1139;
Best Local Similarity 60.7%; Pred. No. 1.4e-120;
Matches 345; Conservative 71; Mismatches 107; Indels 45; Gaps 4;

```

```

QY 266 DIDLKQLYKLPISORHVESDGLVDPDAQITSTRAGVAVPHGDHYHFIPIYSQMSLEEE 325
DB 3 NISLRELYAKPLSEHNVESDGLVDPDAQITSTRAGVAVPHGDHYHFIPIYSQMSLEEK 62
QY 326 RIARIIPLRYSNMWVPSRPEQSPQTPPEPS-----POPAFNPIDKLSNLSVSQL 379
DB 63 RIARIIPLRYSNMWVPSRPEQSPQTPPEPSLOPAFNPIDKLSNLSVSQL 420
QY 380 VRKAGEYVEEKGISRYVPAKDLPSSETVKNLEKSKOSVSHTLTAKKENVAPROEF 439
DB 121 VRKAGDYVEEENGVSRYIIPAKDLISASTAGIDSKLAKQESLSHKLGAKTDLPSDREF 180
QY 440 YDKAYNLLTEAHKALFXNKGNSDFOALDKLERLNDSTNKEKLVDDLAFIAPITHPER 499
DB 181 YDKAYDILARIHODLLNKGROVDFEALDNLRLKVOSSDKVCLVDDILAFIAPIRHPRL 240
QY 500 RLKGNQIETEDVEVIAQLADKYTSDGYIFPEHDIISDEGDAYVTPHMGSHWTKGD 559
DB 241 RLKGNQIETEDVEVIAQLADKYTSDGYIFPEHDIISDEGDAYVTPHMGSHWTKGD 300
QY 560 SLSEKRYVNAQAQYKKEGILPPSPDADYKANPTGDSAAIYNRYKGEKRIPLVLPYVME 619
DB 301 SLSEKRYVNAQAQYKKEGILPPSPDADYKANPTGDSAAIYNRYKGEKRIPLVLPYVME 360
QY 620 HTVEVKNQNLIIIPKDHYNHIFKAMPDDHTYKAPNGYTLDELFTIKYVYEHDPDERPHSN 679
DB 361 HTVEVKNQNLIIIPKDHYNHIFKAMPDDHTYKAPNGYTLDELFTIKYVYEHDPDERPHSN 420
QY 680 DGMGNASEHYLGK-----KXHS 702
DB 421 NGFNASDHVQRNKGADNTQTEKPESEKPOTEKPEEETPREKPOSEKPESEPKPTEEBEE 480
QY 703 FKADPEVEETPAPEVOTEKEVEAQLKEAYLAKVYDSSLKANATEETLAGLRNNLT 762
DB 481 FKADPEVEETPAPEVOTEKEVEAQLKEAYLAKVYDSSLKANATEETLAGLRNNLT 540
QY 763 LQIMDNNSIMAEAEKLLALKGSNPSV 790
DB 541 LQIMDNNSIMAEAEKLLALKESOPAFI 568

```

```

RESULT 9
US-09-884-465A-378
; Sequence 378, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee

```

```

; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 378
; LENGTH: 1378
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Xaa = Methionine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (570)..(570)
; OTHER INFORMATION: Xaa = Glycine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (571)..(571)
; OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-378

```

```

Query Match      43.1%; Score 1795.5; DB 11; Length 1378;
Best Local Similarity 60.7%; Pred. No. 1.9e-120;
Matches 345; Conservative 71; Mismatches 107; Indels 45; Gaps 4;

```

```

QY 266 DIDSLLKQYKPLSQRHVESDGLVFDPAQITSRTARGAVPHGHNHYHPIEYQMSLEK 325
DB 3 NSSLRELIYAKPLSRHVESDGLVFDPAQITSRTARGAVPHGHNHYHPIEYQMSLEK 62
QY 326 RIARIIPLRYSNMHWPPDSRPEQSPQTPPEPS-----PGQPAFLKIDNSSLSVQL 379
DB 63 RIARIIPLRYSNMHWPPDSRPEQSPQTPPEPSQIPAPNPQAPNSNID--EKLVKKA 120
QY 380 VAKVGEYVFEKGISRYFAKDLPSRTVKNLESKLSKQESVSHLTAKKENVAPRDOEF 439
DB 121 VAKVGDGYVFEENGVSRYIPAKDLPSRTVKNLESKLSKQESVSHLTAKKENVAPRDOEF 180
QY 440 YKAVNLITLAKALFXNKGKNSDPQALDKLERLNDESTNKEKLVDDLLAFAPITHE 499
DB 181 YKAVNLITLAKALFXNKGKNSDPQALDKLERLNDESTNKEKLVDDLLAFAPITHE 240
QY 500 RLKGNISQIETDEDEVRIFAQLADKYTSDGYIFDEHDIISDEGDAYVTPHMGSHWIGD 559
DB 241 RLKGNISQIETDEDEVRIFAQLADKYTSDGYIFDEHDIISDEGDAYVTPHMGSHWIGD 300
QY 560 SLSDEKVAQAQYTKKGLIPSPDADVKANPTGSAALYNNRVKGEKRIPLVRLPYWE 619
DB 301 SLSDEKVAQAQYTKKGLIPSPDADVKANPTGSAALYNNRVKGEKRIPLVRLPYWE 360
QY 620 HTVEYKNGULIIPKHQHNHIKFWAMPDHTYKAPNGYITLEDLPATIKYVEHPDERPHN 679
DB 361 HTVEYKNGULIIPKHQHNHIKFWAMPDHTYKAPNGYITLEDLPATIKYVEHPDERPHN 420
QY 680 DGMGVAEHLGK-----KDHS-----DPNKG 702
DB 421 DGMGVAEHLGK-----KDHS-----DPNKG 480
QY 703 FKADEPVEETPAEPEVPQVETEKVEAOLKEAEVLLAKYTDLSLKNATETTLAQRNLT 762
DB 481 FKADEPVEETPAEPEVPQVETEKVEAOLKEAEVLLAKYTDLSLKNATETTLAQRNLT 540
QY 763 LQIMNNISIMAEAKLLALLKGSNPSV 790
DB 541 LQIMNNISIMAEAKLLALLKGSNPSV 568

```

```

RESULT 10
US-09-884-465A-381
; Sequence 381, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biotech, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 381
; LENGTH: 1238
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Xaa = Methionine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (430)..(430)
; OTHER INFORMATION: Xaa = Glycine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (431)..(431)
; OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-381

```

```

Query Match      40.0%; Score 1665; DB 11; Length 1238;
Best Local Similarity 66.5%; Pred. No. 4.2e-111;
Matches 314; Conservative 57; Mismatches 97; Indels 4; Gaps 2;

```

```

QY 266 DIDSLLKQYKPLSQRHVESDGLVFDPAQITSRTARGAVPHGHNHYHPIEYQMSLEK 325
DB 2 DIDSLLKQYKPLSQRHVESDGLVFDPAQITSRTARGAVPHGHNHYHPIEYQMSLEK 61
QY 326 RIARIIPLRYSNMHWPPDSRPEQSPQTPPEPSPPQAPNPQAPNSNID--EKLVKKA 381
DB 62 RIARIIPLRYSNMHWPPDSRPEQSPQTPPEPSPPQAPNPQAPNSNID--EKLVKKA 121
QY 382 KYGGEYVFEKGISRYFAKDLPSRTVKNLESKLSKQESVSHLTAKKENVAPRDOEF 441
DB 122 KYGGEYVFEENGVSRYIPAKDLPSRTVKNLESKLSKQESVSHLTAKKENVAPRDOEF 181
QY 442 KAVNLITLAKALFXNKGKNSDPQALDKLERLNDESTNKEKLVDDLLAFAPITHE 501
DB 182 KAVNLITLAKALFXNKGKNSDPQALDKLERLNDESTNKEKLVDDLLAFAPITHE 241
QY 502 RLKGNISQIETDEDEVRIFAQLADKYTSDGYIFDEHDIISDEGDAYVTPHMGSHWIGD 561
DB 242 RLKGNISQIETDEDEVRIFAQLADKYTSDGYIFDEHDIISDEGDAYVTPHMGSHWIGD 301
QY 562 SLSDEKVAQAQYTKKGLIPSPDADVKANPTGSAALYNNRVKGEKRIPLVRLPYWE 621
DB 302 SLSDEKVAQAQYTKKGLIPSPDADVKANPTGSAALYNNRVKGEKRIPLVRLPYWE 361
QY 622 VEKNGULIIPKHQHNHIKFWAMPDHTYKAPNGYITLEDLPATIKYVEHPDERPHN 681
DB 362 VEKNGULIIPKHQHNHIKFWAMPDHTYKAPNGYITLEDLPATIKYVEHPDERPHN 421
QY 682 WGNASEHVLGKDHSDPNKPKADEPVEETPAEPEVPQVETEKVEAOLKE 733
DB 422 WGNASEHVLGKDHSDPNKPKADEPVEETPAEPEVPQVETEKVEAOLKE 473

```

```

RESULT 11
US-09-884-465A-376
; Sequence 376, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 376
; LENGTH: 999
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Xaa = Methionine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (430)..(430)
; OTHER INFORMATION: Xaa = Glycine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (431)..(431)
; OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-376

Query Match          39.8%; Score 1656; DB 11; Length 999;
Best Local Similarity 61.9%; Pred. No. 1.3e-110;
Matches 326; Conservative 62; Mismatches 99; Indels 40; Gaps 8;

QY 266 DIDLKQLYKLPLOSRHVESDGLVDPDAQITSTRAGVAVPHGDHFIPIYQSMSELE 325
DB 2 DIDLKQLYKLPLOSRHVESDGLVDPDAQITSTRAGVAVPHGDHFIPIYQSMSELEK 61
QY 326 RIARIIPLRYSNNHWPDSRPEPSPOPTPEPSPOPAIPNLK-IDSN---SLVSQLVR 381
DB 62 RIARIIPLRYSNNHWPDSRPEPSPOPTPEPSPOPAIPNLK-IDSN---SLVSQLVR 121
QY 382 KVGEGYVEEKGISRYVFAKDLPSSETVKNLESKLSKQESVSHTLTAKKENVAPRDOEFPD 441
DB 122 KVGEGYVEEKGISRYVFAKDLPSSETVKNLESKLSKQESVSHTLTAKKENVAPRDOEFPD 181
QY 442 KAYNLLTEAHKALFXNKGNSDFOALDKLLERLNDESTNKEKLVDDLLAFAPITHTPERL 501
DB 182 KAYNLLTEAHKALFXNKGNSDFOALDKLLERLNDESTNKEKLVDDLLAFAPITHTPERL 241
QY 502 GKPNQSIETDEVAIAQLADKYTTSDDGIIFDEHDIISDEGDAYVTTPHMGSHWIGKQSL 561
DB 242 GKPNQSIETDEVAIAQLADKYTTSDDGIIFDEHDIISDEGDAYVTTPHMGSHWIGKQSL 301
QY 562 SDKEVAAQAATYKKEGILPPSPDADVANKANPTGDSAAATYNRVKGKRIPLVRLPYMVEHT 621
DB 302 SDKEVAAQAATYKKEGILPPSPDADVANKANPTGDSAAATYNRVKGKRIPLVRLPYMVEHT 361
QY 622 VEKNGNLIIPKDHYNHNIKFAWFDHTYKAPNGYTLLEDLFAITIKYUVEHPDERPHSNDG 681
DB 362 VEKNGNLIIPKDHYNHNIKFAWFDHTYKAPNGYTLLEDLFAITIKYUVEHPDERPHSNDG 421
QY 682 WGNASEHVLGKKHSEDPNKNFKADEBPVEETPA-----EPVPOVETEKVEAQ 730
DB 422 WGNASEHVLGKKHSEDPNKNFKADEBPVEETPA-----EPVPOVETEKVEAQ 730
422 FGNASDHV--XXDLTEE--QIKAAQKHLIEVKTSHNGLSLSHSEODYP-----GN 468

```

```

QY 731 LKAEVLLAKVTDSLSKANATELGLRNNLTLOIMNNSIMAEK 777
DB 469 AKEMKDLDDKI-----EKXIAGIMKQGVK---RESIVVKEK 503

RESULT 12
US-09-884-465A-377
; Sequence 377, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 377
; LENGTH: 999
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: Xaa = Methionine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (570)..(570)
; OTHER INFORMATION: Xaa = Glycine or nothing
; NAME/KEY: MISC_FEATURE
; LOCATION: (571)..(571)
; OTHER INFORMATION: Xaa = Proline or nothing
US-09-884-465A-377

Query Match          39.6%; Score 1649; DB 11; Length 999;
Best Local Similarity 71.7%; Pred. No. 4.3e-110;
Matches 307; Conservative 49; Mismatches 68; Indels 4; Gaps 2;

QY 266 DIDLKQLYKLPLOSRHVESDGLVDPDAQITSTRAGVAVPHGDHFIPIYQSMSELE 325
DB 572 DIDLKQLYKLPLOSRHVESDGLVDPDAQITSTRAGVAVPHGDHFIPIYQSMSELEK 631
QY 326 RIARIIPLRYSNNHWPDSRPEPSPOPTPEPSPOPAIPNLK-IDSN---SLVSQLVR 381
DB 632 RIARIIPLRYSNNHWPDSRPEPSPOPTPEPSPOPAIPNLK-IDSN---SLVSQLVR 691
QY 382 KVGEGYVEEKGISRYVFAKDLPSSETVKNLESKLSKQESVSHTLTAKKENVAPRDOEFPD 441
DB 632 KVGEGYVEEKGISRYVFAKDLPSSETVKNLESKLSKQESVSHTLTAKKENVAPRDOEFPD 691
QY 442 KAYNLLTEAHKALFXNKGNSDFOALDKLLERLNDESTNKEKLVDDLLAFAPITHTPERL 501
DB 692 KAYNLLTEAHKALFXNKGNSDFOALDKLLERLNDESTNKEKLVDDLLAFAPITHTPERL 751
QY 502 GKPNQSIETDEVAIAQLADKYTTSDDGIIFDEHDIISDEGDAYVTTPHMGSHWIGKQSL 561
DB 752 GKPNQSIETDEVAIAQLADKYTTSDDGIIFDEHDIISDEGDAYVTTPHMGSHWIGKQSL 811
QY 562 SDKEVAAQAATYKKEGILPPSPDADVANKANPTGDSAAATYNRVKGKRIPLVRLPYMVEHT 621
DB 812 SDKEVAAQAATYKKEGILPPSPDADVANKANPTGDSAAATYNRVKGKRIPLVRLPYMVEHT 871
QY 622 VEKNGNLIIPKDHYNHNIKFAWFDHTYKAPNGYTLLEDLFAITIKYUVEHPDERPHSNDG 681
DB 872 VEKNGNLIIPKDHYNHNIKFAWFDHTYKAPNGYTLLEDLFAITIKYUVEHPDERPHSNDG 931
QY 932 VEKNGNLIIPKDHYNHNIKFAWFDHTYKAPNGYTLLEDLFAITIKYUVEHPDERPHSNDG 991

```

QY 682 WGNASEHV 689
DB 992 FGNASDHV 999

RESULT 13

US-09-884-465A-6
Sequence 6, Application US/09884465A
Publication No. US2003007723A1
GENERAL INFORMATION:
APPLICANT: Shire Biochem, Inc.
APPLICANT: Hamel, Josee
APPLICANT: Brodeur, Bernard
APPLICANT: Martin, Denis
APPLICANT: Charland, Nathalie
APPLICANT: Ouellet, Catherine
TITLE OF INVENTION: Streptococcus Antigens
FILE REFERENCE: 055190-0044
CURRENT APPLICATION NUMBER: US/09/884,465A
PRIOR FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: 60/212,683
NUMBER OF SEQ ID NOS: 384
SOFTWARE: PatentIn version 3.1
SEQ ID NO 6
LENGTH: 1039
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-884-465A-6

Query Match 29.9%; Score 1247; DB 11; Length 1039;
Best Local Similarity 41.2%; Pred. No. 4,6e-81;
Matches 298; Conservative 95; Mismatches 180; Indels 150; Gaps 19;

QY 1 SYELGIQARTYKVENNRVSYIDGKATOKTENLTPEVSKREGINAEQIVIKITDGGYVT 60
DB 22 AYALNHRQEKNDNRVSYVDSQSOKSENLTPOVSOKEGIOAEQIVIKITDGGYVT 81
QY 61 SHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEDIVNEVKGYYIKYDGKYYVYLKDA 120
DB 82 SHGDHYHYNGKVPYDALFSEELMKDPNYQLKDA DIVNEVKGYYIKYDGKYYVYLKDA 141
QY 121 AAADVNRTEELNRQKOEHSQHRREGTFRNDGAVVALASQGRYTTDDGYIFNADIIEDT 180
DB 142 AAADVNRTEELNRQKOEHSQHRREGTFRNDGAVVALASQGRYTTDDGYIFNADIIEDT 197
QY 181 GDAVYIPHGDDHYHYPKNEISASELAAPFLSGRGNLSNRTYRONSNDTISRNVWVS 240
DB 198 GNAVYIPHGDDHYHYPKNEISASELAAPFLSGRGNLSNRTYRONSNDTISRNVWVS 248
QY 241 VSNPQTNTNTSNNTNSQASQNSNDISLLKQYKLPLOSRRVSDGLVFPDPAQITSR 300
DB 249 -----NTQSVAKGSTSKPANKSENLOSLKELYSPSAQRYSSESDGLVFPDPAKIIISRT 301
QY 301 ARGVAVPHGDHYHFIYSQMSLEERIRARIIPLRYSNHWVPDSRPEQSPQPTPEPSPG 360
DB 302 PNGVAVIPHGDDHYHFIYSKLSALEEKIARMP----- 333
QY 361 POPAPNLKIDSNSLVSQLRVKGEVVEEKGISRYVPAKDLPSSTVKNLESKLSKQES 420
DB 334 -----ISGTGSTVSTNAK-----PNEVVSLSGLSLSNPSS 363
QY 421 VSHHTLAKKENVAPRQDEYDKAYNLLTEAHKALFXNKGNSDFQALDKLELNDESTN 480
DB 364 -----LTTSELSASADGYIFNPK-DIVETATAYIVRHG-DHFHYIPK-----SNQIG 410
QY 481 KEKLVDDLAFAPLITHPERLGKPNQIEYTEDEVARIAQLADKYTTSDGYIFDEHDIISD 540
DB 411 QPTLPNNLSLATPSP-SLPINPGTSHKHE-----EDGCFPANNITIE 452
QY 541 EGDVAYTTPHNGSHMIGKQSLSDKEKYAAQAYTEKGIILPPSPDADYKANFTG-DSAAAI 599
DB 453 DESGFVMSHGDHNYHFFKKDLTEEQILKAQKHLE-----VKTSHNGLDLSLSH 501

QY 600 YNRKVG-----EKRIPLVRLPYMV--EHTVEYKGNLIT-PRKDHYNIKFAM 644
DB 502 EODYPONAKEMKDLDKKIBEKIAGIMKQGVKRESEIVNKEKNAIITPIGDDHHADP---- 558
QY 645 FDDHTYKAPNGYLTLEDLFTATIKYYVEHPDERPHSNDG-----WGNASEHVLGKKHSEDPN 700
DB 559 IDEH---KPVGIG---HSHSNVLELPKPEGVAKKEGNKVYTGEBELTNVNLKNGSTFNN 611
QY 701 KNF 703
DB 612 QNF 614

RESULT 14

US-09-769-787-38
Sequence 38, Application US/09769787
Publication No. US2003009157A1
GENERAL INFORMATION:
APPLICANT: Microbial Technics Limited
APPLICANT: Gilbert, Christophe FG
APPLICANT: Hansbro, Philip M
TITLE OF INVENTION: Proteins
FILE REFERENCE: PWC/P21129WO
CURRENT APPLICATION NUMBER: US/09/769,787
PRIOR FILING DATE: 2001-01-26
PRIOR APPLICATION NUMBER: GB 9816337.1
PRIOR FILING DATE: 1998-03-27
PRIOR APPLICATION NUMBER: US 60/125164
NUMBER OF SEQ ID NOS: 388
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 38
LENGTH: 484
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-769-787-38

Query Match 28.9%; Score 1203.5; DB 11; Length 484;
Best Local Similarity 45.8%; Pred. No. 2e-78;
Matches 259; Conservative 77; Mismatches 120; Indels 109; Gaps 10;

QY 1 SYELGIQARTYKVENNRVSYIDGKATOKTENLTPEVSKREGINAEQIVIKITDGGYVT 60
DB 22 AYALNHRQEKNDNRVSYVDSQSOKSENLTPOVSOKEGIOAEQIVIKITDGGYVT 81
QY 61 SHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEDIVNEVKGYYIKYDGKYYVYLKDA 120
DB 82 SHGDHYHYNGKVPYDALFSEELMKDPNYQLKDA DIVNEVKGYYIKYDGKYYVYLKDA 141
QY 121 AAADVNRTEELNRQKOEHSQHRREGTFRNDGAVVALASQGRYTTDDGYIFNADIIEDT 180
DB 142 AAADVNRTEELNRQKOEHSQHRREGTFRNDGAVVALASQGRYTTDDGYIFNADIIEDT 197
QY 181 GDAVYIPHGDDHYHYPKNEISASELAAPFLSGRGNLSNRTYRONSNDTISRNVWVS 240
DB 198 GNAVYIPHGDDHYHYPKNEISASELAAPFLSGRGNLSNRTYRONSNDTISRNVWVS 248
QY 241 VSNPQTNTNTSNNTNSQASQNSNDISLLKQYKLPLOSRRVSDGLVFPDPAQITSR 300
DB 249 -----NTQSVAKGSTSKPANKSENLOSLKELYSPSAQRYSSESDGLVFPDPAKIIISRT 301
QY 301 ARGVAVPHGDHYHFIYSQMSLEERIRARIIPLRYSNHWVPDSRPEQSPQPTPEPSPG 360
DB 302 PNGVAVIPHGDDHYHFIYSKLSALEEKIARMP----- 333
QY 361 POPAPNLKIDSNSLVSQLRVKGEVVEEKGISRYVPAKDLPSSTVKNLESKLSKQES 420
DB 334 -----ISGTGSTVSTNAK-----PNEVVSLSGLSLSNPSS 363
QY 421 VSHHTLAKKENVAPRQDEYDKAYNLLTEAHKALFXNKGNSDFQALDKLELNDESTN 480
DB 364 -----LTTSELSASADGYIFNPK-DIVETATAYIVRHG-DHFHYIPK-----SNQIG 410

```

QY      481  KERKVDLLFAFLPPIHPHELKGNPQOIEYTDDEVRIQLADKYTSGYFEDSHDIISD 540
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      411  QPTLPNNLSLATPSP-SLPINPGTSHKH-----EDGYFDANRIIAE 452
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      541  EGDAYVTPMHGSHMIGKDSLSPKE 565
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      453  DESGFVMSHGDNHNYFFPKDLTEQ 477
        : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 15
US-09-769

```

US-09-769-744A-24
: Sequence 24, Application US/09769744A
: Publication No. US2003013407A1
: GENERAL INFORMATION:
: APPLICANT: Le Page, Richard WF
: APPLICANT: Wells, Jeremy M
: APPLICANT: Hanniffy, Sean B
: APPLICANT: Hansbro, Philip M
: TITLE OF INVENTION: Proteins
: FILE REFERENCE: PWC/P2112WO
: CURRENT APPLICATION NUMBER: US/09/769, 744A
: PRIORITY FILING DATE: 2001-01-26
: PRIOR APPLICATION NUMBER: PCT/GS99/02452
: PRIOR FILING DATE: 1999-07-27
: PRIOR APPLICATION NUMBER: GB 9816336.3
: PRIOR FILING DATE: 1998-07-27
: PRIOR APPLICATION NUMBER: US 60/125329
: PRIOR FILING DATE: 1999-03-19
: NUMBER OF SEQ ID NOS: 196
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 24
: LENGTH: 484
: TYPE: PRT
: ORGANISM: Streptococcus pneumoniae
US-09-769-744A-24

```

Query Match	28.9%	Score 1203.5;	DB 12;	Length 484;
-------------	-------	---------------	--------	-------------

Qy	1	SYELGTYOATVYKNNRVRVYIDGKQATKTEULTEPEYSKREGIAEIOYIKITDOGYVT	60
Dp	22	AYALNHQRESEKDNKRVSYVDGSSQSKSEMLTDPQYSQKEGIOABOYIKITDOGYVT	81
Qy	61	SHGDHYYNYNGKPYDAITISELLMKDPNYKDKEDIVNEVKGSGYIVKDGKYYVYLKDA	120
Dp	82	SHGDHYYNYNGKPYDALFSEELLMDPNYQYDKADIVNEVKGSGYIIVKDGKYYVYLKDA	141
Qy	121	AHNDNRTXELINROKOEHSOHRREGTTPNDCAVAVLASOGRYTTDDCYIFNABDIIEPT	180
Dp	142	AHNDNRTXDELINROKOEHRKONE---KVNNSVAVASOGRYTTDDGYVNPADIIIEPT	197
Qy	181	GDAYIVPHGDHYHYIRKNELSASELAAAEPLSGRGLNSGRYTRKRONDSNTRTNWPS	240
Dp	198	GNAYIVPHGDHYHYIRKSDLSASELAAAKAHLAGNMQPSQLSYSTASDN-----	248
Qy	241	VSNPGTTNTNTSNKNTNSQASQSNDBLKLQLYKPLSRQHVESDGLVDPQOITSRT	300
Dp	249	-----NQSVAKGSSTKPNKNSSEMLLOSLLELVDPSAQRYSRSDGLVDPFPAKIISRT	301
Qy	301	ARGVAIVPHGDHYHYIRYEQMSSELEERIIAIIPLRYRSHHWPDSSRPEQSPQPRPEPSG	360
Dp	302	PNGVAILPHGDHYHYIRYKLSLALERKIANWV-----	333
Qy	361	POBAPNLKIDSNSLVSQLVKRVGEGYVEEKIGISRYPFANDLPSETYKMLESKLSKQES	420
Dp	334	-----ISGTSITVSTNAK-----PREVYSSLSGTSLSNPSS	363
Qy	421	VSHITLPAKKENVAIPROEFYDRAVNLITFAHQAULPXNKNRSDPQALDKLLERLNDSTN	480
Dp	364	-----LTTSKELSSASGYSYFNPK-DIVETATAYIVRHG--DHFHYIPK-----SNQIG	410

QY 481 KETLVVDLLAFAPAIPIHPERLLKSPQSQIETVEDEVRIMQLADKYTTSDGVI FDEHDIIISD 540
Db 411 QPTLPNNLSATPSP-SLPINPGETSIHKGE-----EDGCGFDMNRITAE 452

RESULT 16
US-09-765

Sequence 182, Application US/09765272
Patent No. US20020061545A1
GENERAL INFORMATION:
APPLICANT: Choi, et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/765,272
FILING DATE: 22-Jan-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/961,083
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 182:
SEQUENCE CHARACTERISTICS:
LENGTH: 447 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 182:
US-09-765-272-182

Query Match	28.5%;	Score 1187.5;	DB 9;	Length 447;
Best Local Similarity	46.3%;	Pred. No. 2.5e-77;		
Matches 257;	Conservative 72;	Mismatches 117;	Indels 109;	Gaps 10

[illegible]

QY 244 PGTNTNTSNNSTNSQASQNSIDSLKQLYKPLISQHVESDGLVFPDPAQITSTARG 303
 Db 225 ---NTQSVAKGSSTSKPANKSENLQSLKELVDSFAQRYSSESDGLVFPDPAKIIISRTPG 280
 QY 304 VAVPHDHHFTFYQMSLEERLARIIPRYSNNHWPDSRPEQSPQPTPEPSGPOP 363
 Db 281 VAIIPHGDHVFIFYSKLSALEEKIARMVP----- 309
 QY 364 APNLKIDSNSLVQSVRVKGEVYEBEKISRYFAKDLPSRTYKNLESKLSKQSVSH 423
 Db 310 -----ISGGSYVSTAK-----PNEVSSLSGLSSNPSS--- 339
 QY 424 TLTAKKENVAPRDOEYDKAYNLLTEAHKALFYKGRNSDPQALDLERLNDESTNKEK 483
 Db 340 -LTSKELSSASDGYIFNPK-DIVEETATAYVRHG--DHFIYIPK-----SNOIGQPT 389
 QY 484 LVDDLLAFAPITTHPERLCKPNSQIYTEDEVRIALADKYTTSDGYIFDEHDIISDEGD 543
 Db 390 LFNNSLATPSP-SLPINPGTSHKHE-----EDGYCPDANRIIAEDES 431
 QY 544 AYVTPHMGSHWIGK 558
 Db 432 GFVMSHGDHNYHFK 446

RESULT 17

US-09-884-465A-384
 ; Sequence 384, Application US/09884465A
 ; Publication No. US2003007293A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shire Biochem, Inc.
 ; APPLICANT: Hamel, Josee
 ; APPLICANT: Brodeur, Bernard
 ; APPLICANT: Martin, Denis
 ; APPLICANT: Charland, Nathalie
 ; APPLICANT: Ouellet, Catherine
 ; TITLE OF INVENTION: Streptococcus Antigens
 ; FILE REFERENCE: 055190-0044
 ; CURRENT APPLICATION NUMBER: US/09/884,465A
 ; CURRENT FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: 60/212,683
 ; PRIOR FILING DATE: 2000-06-20
 ; NUMBER OF SEQ ID NOS: 384
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 384
 ; LENGTH: 913
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Unknown Organism
 ; NAME/KEY: MISC_FEATURE
 ; LOCATION: (1)..(1)
 ; OTHER INFORMATION: Xaa = Methionine or nothing
 ; NAME/KEY: MISC_FEATURE
 ; LOCATION: (344)..(344)
 ; OTHER INFORMATION: Xaa = Glycine or nothing
 ; NAME/KEY: MISC_FEATURE
 ; LOCATION: (345)..(345)
 ; OTHER INFORMATION: Xaa = Proline or nothing
 ; US-09-884-465A-384

Query Match 23.8%; Score 991.5; DB 11; Length 913;
 Best Local Similarity 55.4%; Pred. No. 1e-62;
 Matches 189; Conservative 45; Mismatches 70; Indels 37; Gaps 2;
 QY 487 DLLAFAPITTHPERLCKPNSQIYTEDEVRIALADKYTTSDGYIFDEHDIISDEGDVAV 546
 Db 2 DLLAFAPITTHPERLCKPNSQIYTEDEVRIALADKYTTSDGYIFDEHDIISDEGDVAV 61
 QY 547 TTHMGSHWIGKDSLSDEKKAQAAYTKKGIILPSPDADVKANPFGSDAAIYNNVKCE 606
 Db 62 TTHMGSHWIGKDSLSDEKKAQAAYTKKGIILPSPDADVKANPFGSDAAIYNNVKCA 121

QY 607 KRIPLVRLPYVWEHTVEVNGNLIIPKDHYNIKFAWFDHDTYKAPNGYTLDELJFATIK 666
 Db 122 KVPIDRMVYNNQYIVAVVNGSLIIPHYDHYNIKFEMFDEBLYEAPKYSLEDDLATYK 181
 QY 667 YVVEHPDERPHSNDGNGNASEHYLCK-----KHSE----- 697
 Db 182 YVVEHPNERPHSDNGFGNADHVRKNKADQDSKPEDKHEHDESEPTHPESDEKENHAGL 241
 QY 698 -----DPNKNFADEPVEETPAPEVPQVETKYEQAQKEAEVLLAKYTTDSLTKAN 749
 Db 242 NPSADNLYKPSDTETEETEEADTTDEAETPOVENSVINAKIADABALLEKVTDSIRON 301
 QY 750 ATETLAGLRNNLTLOIMDNNSIMAEKLLALCKSPSSV 790
 Db 302 AMETILGLKSSLLGTCKNNTISAEVDSLALCKESQAPAI 342

RESULT 18

US-09-884-465A-379
 ; Sequence 379, Application US/09884465A
 ; Publication No. US2003007293A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shire Biochem, Inc.
 ; APPLICANT: Hamel, Josee
 ; APPLICANT: Brodeur, Bernard
 ; APPLICANT: Martin, Denis
 ; APPLICANT: Charland, Nathalie
 ; APPLICANT: Ouellet, Catherine
 ; TITLE OF INVENTION: Streptococcus Antigens
 ; FILE REFERENCE: 055190-0044
 ; CURRENT APPLICATION NUMBER: US/09/884,465A
 ; CURRENT FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: 60/212,683
 ; PRIOR FILING DATE: 2000-06-20
 ; NUMBER OF SEQ ID NOS: 384
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 379
 ; LENGTH: 1152
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Unknown Organism
 ; NAME/KEY: MISC_FEATURE
 ; LOCATION: (1)..(1)
 ; OTHER INFORMATION: Xaa = Methionine or nothing
 ; NAME/KEY: MISC_FEATURE
 ; LOCATION: (344)..(344)
 ; OTHER INFORMATION: Xaa = Glycine or nothing
 ; NAME/KEY: MISC_FEATURE
 ; LOCATION: (345)..(345)
 ; OTHER INFORMATION: Xaa = Proline or nothing
 ; US-09-884-465A-379

Query Match 23.8%; Score 991.5; DB 11; Length 1152;
 Best Local Similarity 55.4%; Pred. No. 1.5e-62;
 Matches 189; Conservative 45; Mismatches 70; Indels 37; Gaps 2;
 QY 487 DLLAFAPITTHPERLCKPNSQIYTEDEVRIALADKYTTSDGYIFDEHDIISDEGDVAV 546
 Db 2 DLLAFAPITTHPERLCKPNSQIYTEDEVRIALADKYTTSDGYIFDEHDIISDEGDVAV 61
 QY 547 TTHMGSHWIGKDSLSDEKKAQAAYTKKGIILPSPDADVKANPFGSDAAIYNNVKCE 606
 Db 62 TTHMGSHWIGKDSLSDEKKAQAAYTKKGIILPSPDADVKANPFGSDAAIYNNVKCA 121
 QY 607 KRIPLVRLPYVWEHTVEVNGNLIIPKDHYNIKFAWFDHDTYKAPNGYTLDELJFATIK 666
 Db 122 KVPIDRMVYNNQYIVAVVNGSLIIPHYDHYNIKFEMFDEBLYEAPKYSLEDDLATYK 181
 QY 667 YVVEHPDERPHSNDGNGNASEHYLCK-----KHSE----- 697
 Db 182 YVVEHPNERPHSDNGFGNADHVRKNKADQDSKPEDKHEHDESEPTHPESDEKENHAGL 241

```

Qy 698 -----DPNKNFADEPVEETPEBPQVETEKVEAOIKAEVLLAKVTSLSKAN 749
Db 242 NPSADNLKVPSTDTETEEBAEDTTDEAEIPQVENSVINAKADAEALLEKVTDSISQIN 301
Qy 750 ATETLAGLRNNLTLOIMNNSIMAEAKLLALKKSNSPV 790
Db 302 AMETLTGLKSSLLGCKDNNTISAEVDSLALLKESOPAPI 342

RESULT 19
US-09-884-465A-10
; Sequence 10, Application US/09884465A
; Publication No. US2003007293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; CURRENT FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; PRIOR FILING DATE: 2000-06-20
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 840
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-884-465A-10

```

```

Query Match 23.2%; Score 966.5; DB 11; Length 840;
Best Local Similarity 34.0%; Pred. No. 5.8e-61;
Matches 271; Conservative 100; Mismatches 209; Indels 217; Gaps 28;

Qy 1 SYELGLYQARTVKKNNRVSIDGKQATOKTEVLTPDEVSKEGINABOIVIKITDQGYVT 60
Db 2 AVALNQHSQENKDNRRVSIVGSSQSKSENLTPDOVSQKGIQAEQIVIKITDQGYVT 61
Qy 61 SHGDHYHYNGKVPYDAIISELLMKDPNYKLKDEDIVNEVGKGVIVKDGKYYVYLKDA 120
Db 62 SHGDHYHYNGKVPYDALFSEELMKDPYQOLKQADIVNEVGKGVIIIVKDGKYYVYLKDA 121
Qy 121 AHADVNRKTEINRQKQEHSGHREGGTPRNDGAVALARSGRYTTDDGYIFNADIIEDT 180
Db 122 AHADVNRKTEINRQKQEHVKN-----KVNNSVAVARSQGRYTTNDGVFNPAIDIET 177
Qy 181 GDAYIVPHGDHYHYI PKNELSASELAAEAFLSGRGNLSNSTRYKRONSDNSTRNWPS 240
Db 178 GDAYIVPHGDHYHYI PKSDLSASELAAAVHLAGNMOPQSLSYSTPS-----PS 228
Qy 241 VS-NPGTTNTNTSNNSNTNSQASQSDNIDSLKQLYKLPLSQRHVESDGLVFPDPAQITSR 299
Db 229 LPINGTSH-----EKH-EEDGYGFDAIRIIAE 255
Qy 300 TARQVAVPHGDHYHYI PYSQNSELEERLARIILPRLRSNHVWPDSRPEQSPQPTPEPSP 359
Db 256 DESGFMVSHGDHNYHFFPKDLTEEQIKAAQKLEBVTSHNGLDLSLSEHQDP----- 309
Qy 360 GPQAPNLKIDNSSLSVLQVAKVE--GYVEEKGISGVYVAKQLPSETVKNLESKL-- 415
Db 310 -----SNAEKMDLDRKIEKLAGIMKQGVKR-----ESIVNKEKNAII 350
Qy 416 -----SKOESVSH-----LTAKKENVAPRDOQFVDAKAYILLTEAKALFX 456
Db 351 YPHGDHHAADPIDEHKPVGIGSHSNYELFKPEBGVAKKEG--NKVY---TGEELTNV 404
Qy 457 NGRNSDFQALDKLERLNDESTNKEVLVDLLAFAPITPBERLQKPSQIETVEDEV 516
Db 405 NLKNSTF-----NNQNFLLANGQKRV-----SPSPPELEKKLG----- 439

```

```

Qy 517 IAQLADKYTTSDGYIFDE-----HDIISDEGDAYVTPHMGSHWIGKQSLSDKE- 565
Db 440 -INMLVNLITPDKGVLEKSVSGVFGEGVNIANFELDQPIYLGOTFTYTTIASKQVPEVSY 498
Qy 566 -----KYAAQA--YTKEG-----ILPSPDADVKANPTGDSAAAIYN 601
Db 499 DGFETVPTSLAYKMAQSIIFYPFHAGDLYLRVNPQFAPKGTDALVRFPEFHGNAVLEN 558
Qy 602 RVK-GEKRIPLVRLPYVVEHVEVKNGLIIPHKDHYHNIKF--AWFDHTYKAPRGYT 657
Db 559 NYKVGELKPLPKLNQGTTRIA-----GNKI-----PYTFMANAYLDNOS----- 598
Qy 658 LEDLFATIKYVVEHDERPHSNDGMASEHVLGKQDHSDDPN-----KNFKADE-BPVE 711
Db 599 -----TYIVEP-----ILEKENQTKPSTILPQFKRKAKQENSKLD 634
Qy 712 ETPEPEVPQ-VETEKV 727
Db 635 EKVEPKTSEKVEKEKL 651

```

```

RESULT 20
US-09-769-736-18
; Sequence 18, Application US/09769736
; Publication No. US20030138775A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Le Page, Richard W F
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hamillif, Sean B
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21089WO
; CURRENT APPLICATION NUMBER: US/09/769,736
; CURRENT FILING DATE: 2003-02-14
; PRIOR APPLICATION NUMBER: GB 9816335.5
; PRIOR FILING DATE: 1998-07-27
; PRIOR APPLICATION NUMBER: US 60/125163
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 212
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 822
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae
US-09-769-736-18

```

```

Query Match 22.3%; Score 929; DB 12; Length 822;
Best Local Similarity 29.3%; Pred. No. 2.9e-58;
Matches 270; Conservative 99; Mismatches 236; Indels 318; Gaps 32;

Qy 1 SYELGLYQARTVKKNNRVSID--GKQATOKTENLTPDEVSKEGINABOIVIKITDQGY 57
Db 22 SYOLGKHHMGLATDNDONIAVIDSKGVKAPKT-NKTMDOIISABEGISABEQIVVKITDQGY 80
Qy 58 YVTHSGDHYHYNGKVPYDAIISELLMKDPNYKLKDEDIVNEVGKGVIVKDGKYYVYL 117
Db 81 YVTHSGDHYHYNGKVPYDAIISELLMTDPNHYFKQSDVINEILDDGVIVNNGNYVYL 140
Qy 118 KDAAHADVNRKTEINRQKQEHSGH-BEGG-----TPRNDGAVALARSGRYTTDDGY 169
Db 141 KRGSKRNKIRTKQIQAQVAKGTKEAKKGLAQAVALHSKEEVAVANEAKRGRYTTDDGY 200
Qy 170 IFNADIIEDTGDAYIVPHGDHYHYI PKNELSASELAAEAFLS--GRGLNSRTRYR 226
Db 201 IFSPPTDIIDDGDAYIVPHGDHYHYI PKKDISPELAAQAQVWVSQKQGRG--APPSDVR 258
Qy 227 QNSDNTSRITWVWPVS-NPGTTNTNTSN-----NSNTNSQASQSDN-----IDSLK 272
Db 259 TPAQGRKKAP-IIPVTNPFQGGHOPDNGGYNHAPRPRNDASQNGHQDDEFKGTFFKELLD 317
Qy 273 QLYKLPLSQRHVESDGLVFPDPAQITSRTARQVAVPHGDHYHYI PYSQNSELEERLARIIP 332

```

```

Db 318 HHRLDLKRYHVEDCLIEPTOVISNAFGYVPHGDHYHIIIPRSQSLPELNEMLAD--- 374
Qy 333 LFRYSNHWVPDSRPEQSPQPTPEPSPGPQAPNLKIDNSSLVSQLVKRGEGYVEEK 392
Db 375 -RYLAG-----QTDNDNS-----GSDH----- 390
Qy 393 GISRYFAKDLSEYTKNESKSKOESVSHLTAKENAVAPRDEFYKAYNLLTEAHK 452
Db 391 -----SKRSDKE-VTHFLGHR-----IKAY----- 410
Qy 453 ALFXMKGRNSDPQALDKLERLNDESTNKEKLVDDLAFAPITTPERKGPNSQIETE 512
Db 411 -----GKGLD-----GKP----- 418
Qy 513 DEVRJLAQLADKYTTSBGYIFDEHDILSDEGDAVYTPHMGSHWIGKDSLDEKVAQAAY 572
Db 419 -----YDTSAVYFSGKSHSVKSGVTAGHGFHYITGEGLEQYELDEVANW 467
Qy 573 TKEG---ILPPSPDADYKANFTGDSAAA-----IYNR----- 602
Db 468 VAKAQGADLVAALDQEGKEKPLFTKKVSRKVTQDGKGYIMPKDGKDYFARYQLDL 527
Qy 603 -----VKGER-----IPLVRLPYWEHTVEKKNLIIPIKH 636
Db 528 TQIAFAEQELMKDKKHRYVDIVDTGIEBRLADVSSLPMHAGNATYDTGSSFPIDH 587
Qy 637 YHNIKFAWFDHTYKAPNGYTTLEDLPATIKYVEHDERPHSNDG-----GNASEH 688
Db 588 HHVVPYSWL-----TRNQIATIKYVQHEVVR--DVMSKPGHESSGVITN 632
Qy 689 VLGRKDHSEDPNKNFKADEEPEETPAE-----PE-----VP 720
Db 633 VPLPDKRAGMPWQIISHAEVQKALAEGRFAPDGYIFDPRVLAKETFWMKDGSFSI 692
Qy 721 QVE-----TEKVEAQLKEAVLLAKYDSSLKANATETLAGLRNNLTLOIMDNNSI 771
Db 693 RADGSSLRTINKSDLSQAEWQAOELLAKN---AGDATDT-----DKPE 735
Qy 772 MAEAEKLLALLKGSNPSVSKEK 794
Db 736 KQADK---SNENQOPSEASKEE 755

RESULT 21
; US-09-252-088-15
; Sequence 15, Application US/09252088
; Publication No. US20030031682A1
; GENERAL INFORMATION:
; APPLICANT: BRODEUR, Bernard R.
; APPLICANT: RIOUX, Ciment
; APPLICANT: BOYER, Martine
; APPLICANT: CHARLEBOIS, Isabelle
; APPLICANT: HAMEL, Jose
; APPLICANT: MARTIN, Denis
; TITLE OF INVENTION: NOVEL GROUP B STREPTOCOCCUS ANTIGENS
; FILE REFERENCE: 8331-9002
; CURRENT APPLICATION NUMBER: US/09/252, 088
; EARLIER FILING DATE: 1999-02-18
; EARLIER APPLICATION NUMBER: US/60/075, 425
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 15
; LENGTH: 793
; TYPE: PRT
; ORGANISM: Group B streptococcus
US-09-252-088-15

Query Match 22.3%; Score 927; DB 11; Length 793;
Best Local Similarity 30.7%; Pred. No. 3.8e-58;
Matches 265; Conservative 115; Mismatches 258; Indels 226; Gaps 33;
Qy 14 ENNRVSVID---GKQATOKTENLTPEVSKRGINAQIVIKITDQGYVTSQHDHYVN 70

```

```

Db 6 KNOQIAYIDSSKGAAPKT-NKTMDOISAEBGISAEOIVXKITDQGYVTSQHDHYVN 64
Qy 71 GVPYDAIISEELMKDDPNYKDKEDIVNEVGVIKVDGYVYVYKLOAAHADVRTKE 130
Db 65 GKVPYDAIISEELMTDPPYKQSDVINEIIGGVYKNGVYVYLLKPGSRKRIITQ 124
Qy 131 EINROKESOH-REGG-----TPRNDGAVALARSGRYTTDDGYIFNADIIETDGD 182
Db 125 QIAEQVAKOTKAKERKGLAQAVALHSKEEYAAVNEAKRQRYTTDDGYIFSPDIIIDDLGD 184
Qy 183 AYIVPHGDHYHYIPKNEISASELAAAEFLS---GRGNLSNRYRNRONSNTSRTNWP 239
Db 185 AVLVPNGHNYHYIPKDLSPSELAAQAQVWSQKQRG--ARPSDYRPTPAPRRKAP-IP 241
Qy 240 SVS-NPGTNTNTSN-----NSNTNQASQNSD-----IDSLKLQYKPLSQRYHE 285
Db 242 DVTNPPGQGHQPDNGYHPAPRPNDASQNKHQREFGKTFKELLDOLHRLDLKYRYHE 301
Qy 286 SDGLVFDDPAQITSRTARGAVAPHGDHYHPIPYQKSELEERIA-RILPYRNSHWVPS 344
Db 302 EDGLIFERTQVIXSNAFGYVPHGDHYHIIIPRSQSLPELNEMLADRYLAGQTEDN---DS 357
Qy 345 RPEQSPQPTPEPSPGPAPNLKIDNSSLVSQLVKRGEGYVEEKGIS-RYVFAKDL 403
Db 358 GSEHSKPSDK-----EVTHTFLGHRIKAYGKGLDGKPYDTSADAYFSK-- 400
Qy 404 PSEYTKNESKSKOESVSHLTAKENAVAPRDEFYKAYNLLTEAHKALFXN---KG 459
Db 401 --ESHSVD-----KSGVTAKGDFHYIGFGELEQYELDEVANWVAKAG 443
Qy 460 RNSDFOALDKLERLNDESTNKEKLVDDLAFAPITTPERLG-----KPNQIETEDEV 515
Db 444 Q-----ADELAAL-DQEGKEKPLFTKKVSRKVTQDGKGYIMPKDGKDYFARYQLDL 496
Qy 516 RIAQLADKYTTSBGYIFDEHDILSDEGDAVYTPHMGSHWIGKDSLDEKVAQAAYTKE 575
Db 497 DLTOIA-----FAQELM-----LKDKKHRYVDIV--D 522
Qy 576 KQILPPSPDADYKANFTGDSAAATNRYKGEKRIPLVRLPYWEHTVEKKNLIIPIKH 635
Db 523 TGIEP-----RLAVVSSSLPMHAGNATYDTGSSFPYIPHD 557
Qy 636 YHNIKFAWFDHTYKAPNGYTTLEDLPATIKYVEHDERPHSNDG-----GNASE 687
Db 558 HHVVPYSWL-----TRNQIATIKYVQHEVVR--DVMSKPGHESSGVITN 602
Qy 688 HVLGRKDHSEDPNKNFKADEEPEETPAE-----PE-----V 719
Db 603 NVTPLDKRAGMPWQIISHAEVQKALAEGRFAPDGYIFDPRVLAKETFWMKDGSFSI 662
Qy 720 PQVE-----TEKVEAQLKEAVLLAKYDSSLKANATETLAGLRNNLTLOIMDNNS 770
Db 663 PRADGSSLRTINKSDLSQAEWQAOELLAKN---TGADATDT-----DKPE 705
Qy 771 IMAEAEKLLALLKGSNPSVSKEK 794
Db 706 EKQADK---SNENQOPSEASKEE 726

RESULT 22
; US-09-769-736-24
; Sequence 24, Application US/09769736
; Publication No. US20030138775A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Le Page, Richard WF
; APPLICANT: Wells, Jeremy M
; APPLICANT: Hanniffy, Sean B
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/P21089wo
; CURRENT APPLICATION NUMBER: US/09/769, 736
; CURRENT FILING DATE: 2003-02-14

```

; PRIOR APPLICATION NUMBER: GB 9816335.5
 ; PRIOR FILING DATE: 1998-07-27
 ; PRIOR APPLICATION NUMBER: US 60/125163
 ; PRIOR FILING DATE: 1998-03-19
 ; NUMBER OF SEQ ID NOS: 212
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 24
 ; LENGTH: 381
 ; TYPE: PRT
 ; ORGANISM: Streptococcus agalactiae
 US-09-769-736-24

Query Match 20.2%; Score 841; DB 12; Length 381;
 Best Local Similarity 49.6%; Pred. No. 2e-52;
 Matches 176; Conservative 46; Mismatches 103; Indels 30; Gaps 8;

QY 1 SYELGLYQARTVKEKNNRVSID--GKATQTEMLTPDEVSKREGINAEOIVIKITDQG 57
 DB 22 SYQLGKHHMGLTKDNQIAYIDSXGKVAPKTI-NKTMQISAEIGISAEQIVVKITDQG 80
 QY 58 VYTSGDHYHYNGVYPDAIISELLMKDPYKDKEDIVNEVGKGVYIKYDKGYVYL 117
 DB 81 VYTSGDHYHYNGVYPDAIISELLMKDPYKDKEDIVNEVGKGVYIKYDKGYVYL 140
 QY 118 KDAHADNVRTKEEINROKOEHSOH-REGG-----TPRNDGAVALARSGRYTTDDGY 169
 DB 141 KPGSRKNIIRTKQIAEOVAKGTKEAKGGLAQVAHLSEEVAAVNEAKRGGRYTTDDGY 200
 QY 170 IFNASDIETDGDYIVPHGDHYHYI PKNELSASELAAAEFLS--GRGLNSRTYR 226
 DB 201 IFSPDIIDLDGADLVPHGNHYHYI PKKDLSPSELAAGAYWSQKQGG--ARPSDYR 258
 QY 227 QNSDTSRTNWPSVSNPCTTNTNSN-----NSNTNSQASQND-----IDSLKQ 273
 DB 259 TPAPRRKAPLPDVTNPQGHQPDNGGYHAPRPPNPASQNKQRDEFKGTFFKELLDQ 318
 QY 274 LYKLPLSORHVESDGLVFPDAQITSRTAGVAVPHGDHYHFI PYSQMSLEERIA 328
 DB 319 LHRDLKRYHVEDDLIFEPITQVIXSNAGYVVPBGDHHIIPRQLSLEHMLA 373

RESULT 23
 US-09-884-465A-369
 ; Sequence 369, Application US/09884465A
 ; Publication No. US2003007293A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shire Biochem, Inc.
 ; APPLICANT: Hamel, Josee
 ; APPLICANT: Brodeur, Bernard
 ; APPLICANT: Martin, Denis
 ; APPLICANT: Charland, Nathalie
 ; APPLICANT: Ouellet, Catherine
 ; TITLE OF INVENTION: Streptococcus Antigens
 ; FILE REFERENCE: 055190-0044
 ; CURRENT APPLICATION NUMBER: US/09/884,465A
 ; PRIOR FILING DATE: 2001-06-20
 ; PRIOR APPLICATION NUMBER: 60/212,683
 ; NUMBER OF SEQ ID NOS: 384
 ; SOFTWARE: Patent In version 3.1
 ; SEQ ID NO 369
 ; LENGTH: 906
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Unknown Organism
 US-09-884-465A-369

Query Match 17.4%; Score 723.5; DB 11; Length 906;
 Best Local Similarity 27.3%; Pred. No. 2.2e-43;
 Matches 262; Conservative 105; Mismatches 290; Indels 303; Gaps 38;
 QY 43 GINAGQIVIKITDQGVVSHGDHYHYNGKVYPDAIISELLMKDPYKDKEDIVNEVK 102

DB 37 GFDNRITAE-DEGFVWSHSDSNHYF-----FKXD-----LTREQI KAAQK 77
 QY 103 GGVYIKYDKGYVYLLKDAHADNVRTKEEINROKOEHSQHEGGTPR-----NDGAV 154
 DB 78 -----HLEEVKTSNGDLSISSHEQDYPGNAKENDLDDKIEEKIA 118
 QY 155 ALARSQGRYTTDDGYIFNASDIIEDTGDYIVPHGDHYHYI PKNELSASELAA-----EA 210
 DB 119 GIMKQYG-----VKRESIVNKKERKAI IYPSGDHHAADIPDEHKPVGIGHSHSNEL 170
 QY 211 FLSGRG-----NLSNSTRYRQN-----SDNTSRITNW----- 237
 DB 171 FKPEGVAKKEGKNKYTGEEELTNVNNLLKNSTFNNQFTLLANGQKRVSPPELEKKLG 230
 QY 238 -----VPSVSNCTNTNTNSN-----NSNTNSQASQNDIDSLKQLYKPLSOR---HVESD 287
 DB 231 INMLVKLITPPGKYLEKVGKVFGEVGNINANFELDDPYLFGQTFKYTTASKDYPVSYD 290
 QY 288 GLVFPDAQITSRTAGV-----AVPHG-----DHYHFI PYSQ-- 319
 DB 291 GTFTVPTSLAYKMASQITIFPFHAGDTYLRVNPQPAVPKGTDALVRYVDEFGHAYLENN 350
 QY 320 --MSELEERTARI-----IPLRYSNHWVPDSR-----PEQSPQPTPEPSG 360
 DB 351 KYGGEIKLPKLNQGTTRTAGNKIPVTFMANAVLDNGSTYIVBVLPEKENQTDKPSIL 410
 QY 361 PQ-----PAPNLKD-----SNSL-----VQOLRVK 383
 DB 411 PQFRKNRAQOEKSKDEKVEBKTSKEVEKEKLSFGTGNSTNSLSEVPYTPDPOEKAKF 470
 QY 384 GEGVYFEEKGSRVYPAKD-----LPS-ETVK-----NLESLSKQESVSH 423
 DB 471 AESTGMR---LENVLFPMNDGTIELYLPBGEVYIKOMADFTGEARQNGENKPSGKYS- 526
 QY 424 TLTAKKENVAPRDEPFYDKAYNLLTEA-----HKALFXNKGKNSDFQALDKL 471
 DB 527 --TGTVEN-----QPTENKPADSLPEAPNEKRVKPENSTDGMNPGCNVGS DMLDPAL 579
 QY 472 ERLNDESTNKKLYD-----DLAFLAPITHPERLG-----KPR-----SQLEY 510
 DB 580 EEAPAVDPVQOKLEKFTASVGLGDSVIFNNMDGTIELRLPBGEVYIKNLSDFIAGPQITY 639
 QY 511 TEDEVRIAQADKYTTSDGYIFDEHDIISDEGDAYVTPPHMGSNMGKSDSDXEKVAQ 570
 DB 640 TDEIQUAKLAGKXTTEDGYIFD-----TSWIKDSDSEBRRAAQ 680
 QY 571 AYTEKGLPPSPDADYKANPTGDSAAIYVRVGEKRIPLVRLPYVVEHTVEYKNGULI 630
 DB 681 AYAKEKGLTPPSTHDQSGNTEAKGAEAIVYRKAACKVPLDRMPYNLQYTVVEYNGSLI 740
 QY 631 IPHKDHYNIKFANFPDDHTTYAAPNGYTLDELFAITIKYVVEIPDRPSNDQMGASERV- 689
 DB 741 IPHYDHYNIKFEMFPDEGLYEAAPKGYSLDELATVXYVE---PR-----NASDVR 789
 QY 690 LGKXKHSEDPKKNKADE--EVEETPAEPE-----VPQVETEKEVAOLK-- 732
 DB 790 KKKADQOSKPEDEKHEVSEPTTPESDDEKKNHAGLNPADNLKYPSTDTETETEBEADT 849
 QY 733 --EAEVLLAKVYTDSSLANATETIAGLRNNLITLOIMDNNSIMAEAKLLALLKSNPSSV 790
 DB 850 TDEAEI-----PGTFSINQNAMETILTGLKSSLLTCKONNTISAVVDSLALAKESQAPAI 905

RESULT 24
 US-09-884-465A-373
 ; Sequence 373, Application US/09884465A
 ; Publication No. US2003007293A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Shire Biochem, Inc.
 ; APPLICANT: Hamel, Josee
 ; APPLICANT: Brodeur, Bernard
 ; APPLICANT: Martin, Denis


```
Db 43 AAAQAYAKEGILTPSPSTHDQSGNTEAKGAELIYNRVKAAKVPIDRMPYNIQYTVVEYKN 102
Qy 627 GNLIIIPKHDYHNIFKAFMDHTYKAPNGYTLDELPAITIKYVVEHPDRPHSNDGNGAS 686
Db 103 GSLIIPHDSTHNIFKFEWFDGLYEPKGYSLDELPAITIKYVE-----PR-----NAS 151
Qy 687 EHV-LGKKDHSEDPKNFKADE--EPVEETPAEP-----VPOVETEKVBAQ 730
Db 152 DHVRKNKADQDSKPEDKEHDEVSEPTHPESDEKENHAGLNPASADNLKYPSTDETEEBE 211
Qy 731 LK---EAEVLLAKYTDSSLKANATETLAGLRNNLTLOIMDNNSIMAEAEKLLALLKGSN 786
Db 212 AEDTDEAEI-----PGTPIRQNAEMETLTGLKSSLLGTOKNNITSAEVDLSLLALKESQ 267
Qy 787 PSSV 790
Db 268 PAPI 271

RESULT 40
US-09-884-465A-295
; Sequence 295, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 295
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-295

Query Match 15.2%; Score 634; DB 11; Length 272;
Best Local Similarity 45.7%; Pred. No. 1e-37;
Matches 139; Conservative 39; Mismatches 72; Indels 54; Gaps 8;

Qy 507 QIETDEEVRIAGLADKYTTSIDGYIFDEHDIISDEGDAVYTPHMGHSWICKDSLDEK 566
Db 2 QITTYDDEIOYAKLAGKTYTDEGYIFD-----TSWIKDSLSEKER 42
Qy 567 VAAQAYTEKEGILTPSPDADYKANPTGDSAAAIYNRVKGEKRIPLVRLPYWEHTVEYKN 626
Db 43 AAAQAYAKEGILTPSPSTHDQSGNTEAKGAELIYNRVKAAKVPIDRMPYNIQYTVVEYKN 102
Qy 627 GNLIIIPKHDYHNIFKAFMDHTYKAPNGYTLDELPAITIKYVVEHPDRPHSNDGNGAS 686
Db 103 GSLIIPHDSTHNIFKFEWFDGLYEPKGYSLDELPAITIKYVE-----PR-----NAS 151
Qy 687 EHV-LGKKDHSEDPKNFKADE--EPVEETPAEP-----VPOVETEKVBAQ 730
Db 152 DHVRKNKADQDSKPEDKEHDEVSEPTHPESDEKENHAGLNPASADNLKYPSTDETEEBE 211
Qy 731 LK---EAEVLLAKYTDSSLKANATETLAGLRNNLTLOIMDNNSIMAEAEKLLALLKGSN 786
Db 212 AEDTDEAEI-----PGTPIRQNAEMETLTGLKSSLLGTOKNNITSAEVDLSLLALKESQ 267
Qy 787 PSSV 790
Db 268 PAPI 271
```

```
RESULT 41
US-09-884-465A-340
; Sequence 340, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: 60/212,683
; NUMBER OF SEQ ID NOS: 384
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 340
; LENGTH: 894
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Unknown Organism
US-09-884-465A-340

Query Match 15.1%; Score 627; DB 11; Length 894;
Best Local Similarity 45.0%; Pred. No. 2e-36;
Matches 139; Conservative 40; Mismatches 76; Indels 54; Gaps 8;

Qy 507 QIETDEEVRIAGLADKYTTSIDGYIFDEHDIISDEGDAVYTPHMGHSWICKDSLDEK 566
Db 2 QITTYDDEIOYAKLAGKTYTDEGYIFD-----TSWIKDSLSEKER 42
Qy 567 VAAQAYTEKEGILTPSPDADYKANPTGDSAAAIYNRVKGEKRIPLVRLPYWEHTVEYKN 626
Db 43 AAAQAYAKEGILTPSPSTHDQSGNTEAKGAELIYNRVKAAKVPIDRMPYNIQYTVVEYKN 102
Qy 627 GNLIIIPKHDYHNIFKAFMDHTYKAPNGYTLDELPAITIKYVVEHPDRPHSNDGNGAS 686
Db 103 GSLIIPHDSTHNIFKFEWFDGLYEPKGYSLDELPAITIKYVE-----PR-----NAS 151
Qy 687 EHV-LGKKDHSEDPKNFKADE--EPVEETPAEP-----VPOVETEKVBAQ 730
Db 152 DHVRKNKADQDSKPEDKEHDEVSEPTHPESDEKENHAGLNPASADNLKYPSTDETEEBE 211
Qy 731 LK---EAEVLLAKYTDSSLKANATETLAGLRNNLTLOIMDNNSIMAEAEKLLALLKGSN 786
Db 212 AEDTDEAEI-----PGTPIRQNAEMETLTGLKSSLLGTOKNNITSAEVDLSLLALKESQ 267
Qy 787 PSSVSEKI 795
Db 268 PAPIGPDI 276

RESULT 42
US-09-884-465A-342
; Sequence 342, Application US/09884465A
; Publication No. US20030077293A1
; GENERAL INFORMATION:
; APPLICANT: Shire Biochem, Inc.
; APPLICANT: Hamel, Josee
; APPLICANT: Brodeur, Bernard
; APPLICANT: Martin, Denis
; APPLICANT: Charland, Nathalie
; APPLICANT: Ouellet, Catherine
; TITLE OF INVENTION: Streptococcus Antigens
; FILE REFERENCE: 055190-0044
; CURRENT APPLICATION NUMBER: US/09/884,465A
; PRIOR FILING DATE: 2001-06-20
```


Db 212 AEDTDEABI----PGTSPIRONAMETLTGLKSSLLGTCKONNTISAEDVSLALLKESQ 267
 QY 787 PSSVSEKEI 795
 Db 268 PAPIQGPQI 276

RESULT 45 US-09-884-465A-302

Sequence 302, Application US/09884465A
 Publication No. US20030077293A1
 GENERAL INFORMATION:
 APPLICANT: Shire Biochem, Inc.
 APPLICANT: Hamel, Josee
 APPLICANT: Brodeur, Bernard
 APPLICANT: Martin, Denis
 APPLICANT: Charland, Nathalie
 APPLICANT: Ouellet, Catherine
 TITLE OF INVENTION: Streptococcus Antigens
 FILE REFERENCE: 055190-0044
 CURRENT APPLICATION NUMBER: US/09/884,465A
 PRIOR FILING DATE: 2001-06-20
 PRIOR FILING DATE: 2000-06-20
 NUMBER OF SEQ ID NOS: 384
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 302
 LENGTH: 270
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Unknown Organism
 US-09-884-465A-302

Query Match 15.0%; Score 626; DB 11; Length 270;
 Best Local Similarity 45.7%; Pred. No. 3,8e-37;
 Matches 139; Conservative 39; Mismatches 70; Indels 56; Gaps 9;

QY 507 QIETEDBVRIAQLADKTTSDGYTFDEHDIISDEGDVYVTPHMGSHWICKDSLSDEK 566
 Db 2 QITYTDEIQVAKLAGKTYTDEGYTFD-----TSWIKKDSISEAER 42
 QY 567 VAAQAYTKEKGIPLPSPDADVKANPTGDSAAAIYRVKGEKRIPLVRLPYMVEHTVEYKN 626
 Db 43 AAAQAYTKEKGIPLPSTHDGSGNTEAKGAELIYRVKAAKVPIDRPNYLOYTEVEYKN 102
 QY 627 GNLIIIPKHDYHNKIFAMFDDHTYKAPNGYTLDELFAITIKYVEHPDERPHSNDGWNAS 686
 Db 103 GSLIIP--DHYHNKIFEFWDEGLYAPKGYSLDELATVKKYVE-----PR-----NAS 149
 QY 687 EHV-LGKKDSEDPKNKPKADE--EPVEETPAEP-----VPOVETKEVQAQ 730
 Db 150 DHVRKNKADQSKPEDKEHDEVSPTHPESDEKENHAGLNPADNLKPFSTDEETEE 209
 QY 731 LK----BAEVLAKVTDSSLKANATETLAGLRNNLTLOIMDNNSIMAEAKLLALLKGSN 786
 Db 210 AEDTDEABI----PGTSPIRONAMETLTGLKSSLLGTCKONNTISAEDVSLALLKESQ 265
 QY 787 PSSV 790
 Db 266 PAPI 269

RESULT 46
 US-09-884-465A-296
 Sequence 296, Application US/09884465A
 Publication No. US20030077293A1
 GENERAL INFORMATION:
 APPLICANT: Shire Biochem, Inc.
 APPLICANT: Hamel, Josee
 APPLICANT: Brodeur, Bernard
 APPLICANT: Martin, Denis
 APPLICANT: Charland, Nathalie

APPLICANT: Ouellet, Catherine
 TITLE OF INVENTION: Streptococcus Antigens
 FILE REFERENCE: 055190-0044
 CURRENT APPLICATION NUMBER: US/09/884,465A
 PRIOR FILING DATE: 2001-06-20
 PRIOR APPLICATION NUMBER: 60/212,683
 PRIOR FILING DATE: 2000-06-20
 NUMBER OF SEQ ID NOS: 384
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 296
 LENGTH: 272
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Unknown Organism
 US-09-884-465A-296

Query Match 15.0%; Score 625; DB 11; Length 272;
 Best Local Similarity 45.4%; Pred. No. 4.6e-37;
 Matches 138; Conservative 39; Mismatches 73; Indels 54; Gaps 8;

QY 507 QIETEDBVRIAQLADKTTSDGYTFDEHDIISDEGDVYVTPHMGSHWICKDSLSDEK 566
 Db 2 QITYTDEIQVAKLAGKTYTDEGYTFD-----TSWIKKDSISEAER 42
 QY 567 VAAQAYTKEKGIPLPSPDADVKANPTGDSAAAIYRVKGEKRIPLVRLPYMVEHTVEYKN 626
 Db 43 AAAQAYTKEKGIPLPSTHDGSGNTEAKGAELIYRVKAAKVPIDRPNYLOYTEVEYKN 102
 QY 627 GNLIIIPKHDYHNKIFAMFDDHTYKAPNGYTLDELFAITIKYVEHPDERPHSNDGWNAS 686
 Db 103 GSLIIPSDSYHNKIFEFWDEGLYAPKGYSLDELATVKKYVE-----PR-----NAS 151
 QY 687 EHV-LGKKDSEDPKNKPKADE--EPVEETPAEP-----VPOVETKEVQAQ 730
 Db 152 DHVRKNKADQSKPEDKEHDEVSPTHPESDEKENHAGLNPADNLKPFSTDEETEE 211
 QY 731 LK----BAEVLAKVTDSSLKANATETLAGLRNNLTLOIMDNNSIMAEAKLLALLKGSN 786
 Db 212 AEDTDEABI----PGTSPIRONAMETLTGLKSSLLGTCKONNTISAEDVSLALLKESQ 267
 QY 787 PSSV 790
 Db 268 PAPI 271

RESULT 47
 US-09-884-465A-298
 Sequence 298, Application US/09884465A
 Publication No. US20030077293A1
 GENERAL INFORMATION:
 APPLICANT: Shire Biochem, Inc.
 APPLICANT: Hamel, Josee
 APPLICANT: Brodeur, Bernard
 APPLICANT: Martin, Denis
 APPLICANT: Charland, Nathalie
 APPLICANT: Ouellet, Catherine
 TITLE OF INVENTION: Streptococcus Antigens
 FILE REFERENCE: 055190-0044
 CURRENT APPLICATION NUMBER: US/09/884,465A
 PRIOR FILING DATE: 2001-06-20
 PRIOR APPLICATION NUMBER: 60/212,683
 PRIOR FILING DATE: 2000-06-20
 NUMBER OF SEQ ID NOS: 384
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 298
 LENGTH: 272
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Unknown Organism
 US-09-884-465A-298


```

1  APPLICANT: Shire Biochem, Inc.
2  APPLICANT: Hamel, Joseph
3  APPLICANT: Brodeur, Bernard
4  APPLICANT: Martin, Denis
5  APPLICANT: Charland, Nathalie
6  APPLICANT: Ouellet, Catherine
7  TITLE OF INVENTION: Streptococcus Antigens
8  FILE REFERENCE: 055190-0044
9  CURRENT APPLICATION NUMBER: US/09/884,465A
10 CURRENT FILING DATE: 2001-06-20
11 PRIOR APPLICATION NUMBER: 60/212,683
12 PRIOR FILING DATE: 2000-06-20
13 NUMBER OF SEQ ID NOS: 384
14 SOFTWARE: PatentIn version 3.1
15 SEQ ID NO 345
16 LENGTH: 901
17 TYPE: PRN
18 ORGANISM: Artificial Sequence
19 FEATURE:
20 OTHER INFORMATION: Unknown Organism
21 US-09-884,465A

```

Query Match	15.0%;	Score 625;	DB 11;	Length 901;
Best Local Similarity	45.4%;	Pred. No. 2.8e-36;		
Matches 138;	Conservative 39;	Mismatches 73;	Indels 54;	Gaps 8;

```

Oy 507 QIEYEDERIRQOLDKRYTSSDGYIFDEHDIISDESDAVYTTPMHGSHMIGKSLSDKEK 566
Db 2 QITTYDDDELQVAKLAKKTYTTEDDGYIYD-----TSNKKKSLSEAEK 42
Oy 567 VAAQAYTEKGIPLPSPEADAVKANPTGDSAAAIYNNKVBKBRJPLVRLPYMEVHEVKN 626
Db 43 AAAQYAAKEKGIPLPSTPDHQDSGNETAKAEALYNNVVKAKKPLRMPYNIQYTAEVKN 1020
Oy 627 GNLIIPHXKHYNIRPAMFEDDHTYKAPNGYLTLEDLFATIKYUYEHDERPHSNDGNGNS 686
Db 103 GSLIIPSYDSYNIKFEWDEDEGYEAPKGYSLLEDLATYKYVE-----PR-----MAS 151
Oy 687 EHV-LGKKDHSDEPNKNFKADE--EPEVEETPAEPE-----VPQVETEKVEAO 730
Db 152 DHVRKNKKDDQSKPREDKEHDEVESEETPEDESKEKNHAGINPSADNLYKPSSTDEETEE 211
Oy 731 LK-----EAAVLLAKYTDSSLKANATYTLGLRNNTLQIIMDNNSIIMAAEKLTALIKGSN 786
Db 212 AEDTTEDEAEI-----FOTPSIRONAMETTLGLKSSLLGTCKDNNTTISAEVDSLALLAKESQ 267
Oy 787 PSSV 790
Db 268 PAPI 271

```

Search completed: November 14, 2003, 10:50:45
Job time : 343 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 14, 2003, 01:12:43 ; Search time 8745 Seconds
(without alignments)
11175.883 Million cell updates/sec

Title: US-09-765-271-55
2389

Perfect score: 1 TTCTTACGAGTTGGAGCTGT.....TAAGTAAGGAAAAATAAAC 2389
Sequence:

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 2045481386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 200 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_cm:*
5: gb_ov:*
6: gb_pac:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pac:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pin:*
35: em_hcg_rtd:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_sy:*
39: em_hcgo_hum:*
40: em_hcgo_mus:*
41: em_hcgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2389	100.0	2389	6	ARI20265 Sequence
2	2389	100.0	2389	6	BD063374 Streptococ
3	2389	100.0	2541	1	AF291695 Streptococ
4	2338	97.9	2406	6	AX569139 Sequence
5	2338	97.9	8195	6	AR218862 Sequence
6	2338	97.9	8195	6	BD003774 Polynucle
7	2338	97.9	10256	1	AE007418 Streptococ
8	2338	97.9	349980	6	AX571763 Sequence
9	1524	68.0	10320	1	AE008479 Streptococ
10	927	38.8	20035	2	SPNEU1915 Streptococ
11	326	13.6	702	2	SPNEU1929 Streptococ
12	200	8.4	2535	1	AF340221 Streptococ
13	148	6.2	2523	6	AX343072 Sequence
14	148	6.2	2647	6	AX343073 Sequence
15	53	2.2	2166	12	AF340222 Synthetic
16	53	2.2	2290	6	ARI20270 Sequence
17	53	2.2	2290	6	BD063279 Streptococ
18	53	2.2	2359	6	AR219011 Sequence
19	53	2.2	2359	6	BD003923 Polynucle
20	53	2.2	2457	1	AF318954 Streptococ
21	53	2.2	2457	6	AX569137 Sequence
22	53	2.2	2517	1	AF318955 Streptococ
23	53	2.2	2639	6	AX568778 Sequence
24	53	2.2	2639	6	AX343074 Sequence
25	53	2.2	11931	1	AE007403 Streptococ
26	53	2.2	12372	1	AE009464 Streptococ
27	53	2.2	75874	2	SPNEU1907 Streptococ
28	53	2.2	232807	2	SPNEU1901 Streptococ
29	53	2.2	349980	6	AX571762 Sequence
30	50	2.1	492	12	AF340223 Synthetic
31	44	1.8	973	6	ARI219123 Sequence
32	44	1.8	973	6	BD004035 Polynucle
33	30	1.3	40	6	ARI20406 Sequence
34	30	1.3	40	6	BD063415 Streptococ
35	27	1.1	37	6	ARI20405 Sequence
36	27	1.1	37	6	BD063414 Streptococ
37	25	1.0	33	6	AX343114 Sequence
38	25	1.0	34	6	AX343116 Sequence
39	24	1.0	1342	6	ARI20328 Sequence
40	24	1.0	1342	6	BD063337 Streptococ
41	24	1.0	1446	6	AX608407 Sequence
42	24	1.0	2528	6	AX343078 Sequence
43	24	1.0	3117	6	AX568780 Sequence
44	24	1.0	3120	1	AF318956 Streptococ
45	24	1.0	3120	6	AX343070 Sequence
46	24	1.0	5048	6	AX343071 Sequence
47	24	1.0	6867	6	AR218960 Sequence
48	24	1.0	6867	6	BD003872 Polynucle
49	24	1.0	21397	1	AE014279 Streptococ
50	24	1.0	75248	6	AX602204 Sequence
51	24	1.0	98050	1	SAG766854 Streptococ
52	23	1.0	154101	2	AC136885 Sus scrofa
53	23	1.0	194474	2	AC137536 Sus scrofa
54	23	1.0	203434	10	AL732478 Mouse DNA
55	23	1.0	214633	2	AC115959 Mus muscu
56	22	0.9	1450	8	BT005616 Arabidops
57	22	0.9	1547	8	AY089131 Arabidops
58	22	0.9	1603	8	BT004286 Arabidops
59	22	0.9	54419	2	AC015298 Drosophil
60	22	0.9	103534	8	ATTS58 Arabidops
61	22	0.9	168601	3	AC123445 Rattus no
62	22	0.9	171569	2	AC007811 Drosophil
63	22	0.9	187517	9	AF225898 Homo sapi
64	22	0.9	198084	9	AF225898 Homo sapi
65	22	0.9	212941	3	AE003719 Drosophil

66	22	0.9	221952	2	AC098902	AC098902 Rattus no	c 139	20	0.8	106689	9	AC008919	AC008919 Homo sapi
67	22	0.9	221952	2	AC098902	AC098902 Rattus no	140	20	0.8	110000	2	PFMAL7p1_02	Continuation (3 of
68	22	0.9	231485	2	AC098427	AC098427 Rattus no	141	20	0.8	110000	2	PFMAL7p1_03	Continuation (4 of
69	21	0.9	29	6	AX343113	AX343113 Sequence	142	20	0.8	110590	2	EX001021	EX001021 Dario rer
70	21	0.9	222	14	STVM7004AE	M96201 Stmian immu	143	20	0.8	119995	10	AL731648	AL731648 Mouse DNA
71	21	0.9	2065	8	AF406703	AF406703 Solanum t	144	20	0.8	127313	2	AC078977	AC078977 Oryza sat
72	21	0.9	29488	2	AC014153	AC014153 Drosophila	145	20	0.8	131316	2	AC031985	AC031985 Homo sapi
73	21	0.9	58408	2	EX510940	EX510940 Dario rer	146	20	0.8	133418	2	AL353709	AL353709 Human DNA
74	21	0.9	70046	2	AC120849	AC120849 Mus muscu	147	20	0.8	137173	2	AC016631	AC016631 Homo sapi
75	21	0.9	70046	2	AC120849	AC120849 Mus muscu	148	20	0.8	137519	8	ATT1P17	ATT1P17 Arabidops
76	21	0.9	129559	2	AL731563	AL731563 Human DNA	149	20	0.8	137586	9	AC106775	AC106775 Homo sapi
77	21	0.9	133337	9	AC005550	AC005550 Homo sapi	150	20	0.8	144722	2	AC008831	AC008831 Homo sapi
78	21	0.9	139357	9	AC010907	AC010907 Homo sapi	151	20	0.8	147829	2	AC068806	AC068806 Mus muscu
79	21	0.9	146690	2	AC102254	AC102254 Mus muscu	152	20	0.8	148592	2	AC125190	AC125190 Mus muscu
80	21	0.9	148348	2	AC104074	AC104074 Homo sapi	153	20	0.8	150192	2	AC122555	AC122555 Mus muscu
81	21	0.9	150972	9	AL512380	AL512380 Human DNA	154	20	0.8	151857	2	AL831770	AL831770 Human DNA
82	21	0.9	156105	9	AC020977	AC020977 Homo sapi	155	20	0.8	152806	2	AC129725	AC129725 Rattus no
83	21	0.9	156526	3	AC010665	AC010665 Drosophila	156	20	0.8	156724	2	AC016306	AC016306 Homo sapi
84	21	0.9	158519	9	AL135926	AL135926 Human DNA	157	20	0.8	157812	2	AC055805	AC055805 Homo sapi
85	21	0.9	168544	9	AL513307	AL513307 Human DNA	158	20	0.8	157850	9	AC034303	AC034303 Homo sapi
86	21	0.9	177385	9	AC010885	AC010885 Homo sapi	159	20	0.8	161279	2	AC125902	AC125902 Rattus no
87	21	0.9	180629	2	AC104075	AC104075 Homo sapi	160	20	0.8	161311	2	AC137012	AC137012 Rattus no
88	21	0.9	180643	2	AC117257	AC117257 Mus muscu	161	20	0.8	163709	2	AC132473	AC132473 Mus muscu
89	21	0.9	181547	2	AC108815	AC108815 Mus muscu	162	20	0.8	166226	5	AL929548	AL929548 Homo sapi
90	21	0.9	184365	9	HS109M15	HS109M15 Homo sapi	163	20	0.8	168842	5	AL929548	AL929548 Zebrafish
91	21	0.9	188307	2	AC120080	AC120080 Rattus no	164	20	0.8	170167	10	AL672143	AL672143 Mouse DNA
92	21	0.9	191750	2	EX517317	EX517317 Dario rer	165	20	0.8	171115	9	AC068385	AC068385 Homo sapi
93	21	0.9	193433	2	AP001523	AP001523 Homo sapi	166	20	0.8	171370	9	AC004021	AC004021 Human PAC
94	21	0.9	193547	9	AC024569	AC024569 Homo sapi	167	20	0.8	174325	2	AC136604	AC136604 Homo sapi
95	21	0.9	207698	9	AC114491	AC114491 Homo sapi	168	20	0.8	175838	2	AC116356	AC116356 Homo sapi
96	21	0.9	219687	2	AC119573	AC119573 Mus muscu	169	20	0.8	177071	2	AC128090	AC128090 Rattus no
97	21	0.9	223626	2	AC098602	AC098602 Rattus no	170	20	0.8	177623	2	AC067825	AC067825 Homo sapi
98	21	0.9	226427	2	AC106341	AC106341 Rattus no	171	20	0.8	180450	3	AE014835	AE014835 Plasmodiu
99	21	0.9	236764	2	AC112436	AC112436 Rattus no	172	20	0.8	180571	2	AC113425	AC113425 Homo sapi
100	21	0.9	240929	2	AC130777	AC130777 Rattus no	173	20	0.8	181584	2	AC017027	AC017027 Homo sapi
101	21	0.9	258290	2	AC095878	AC095878 Rattus no	174	20	0.8	182446	2	AC100500	AC100500 Mus muscu
102	21	0.9	291061	2	AC099169	AC099169 Rattus no	175	20	0.8	182632	9	AC007282	AC007282 Homo sapi
103	21	0.9	302417	3	AE003482	AE003482 Drosophila	176	20	0.8	182658	2	AC084743	AC084743 Mus muscu
104	21	0.9	349980	6	AX590980	AX590980 Sequence	177	20	0.8	185845	9	AC121248	AC121248 Homo sapi
105	20	0.8	406	11	GS0086	GS0086 SHGC-82980	178	20	0.8	186165	2	AC112187	AC112187 Homo sapi
106	20	0.8	501	6	AX300897	AX300897 Sequence	179	20	0.8	189350	2	AC136628	AC136628 Homo sapi
107	20	0.8	670	6	AX602072	AX602072 Sequence	180	20	0.8	191231	10	AC126277	AC126277 Mus muscu
108	20	0.8	1910	1	SAG290952	AB030952 Streptococ	181	20	0.8	191553	2	AL772233	AL772233 Rattus no
109	20	0.8	2313	1	AB031552	AB031552 Bacillus	182	20	0.8	192096	2	AC126838	AC126838 Rattus no
110	20	0.8	2469	6	AX088376	AX088376 Sequence	183	20	0.8	192129	2	AC134667	AC134667 Mus muscu
111	20	0.8	2469	6	AX607520	AX607520 Sequence	184	20	0.8	192410	2	AC135831	AC135831 Rattus no
112	20	0.8	2478	1	AB073859	AB073859 Streptococ	185	20	0.8	193057	9	AC022441	AC022441 Homo sapi
113	20	0.8	2478	6	AX088372	AX088372 Sequence	186	20	0.8	193301	9	CNS01DRJ	AL117259 Human chr
114	20	0.8	3501	1	AF062533	AF062533 Streptococ	187	20	0.8	194425	5	AL807818	AL807818 Zebrafish
115	20	0.8	5158	6	AX300905	AX300905 Sequence	188	20	0.8	194907	2	AC136047	AC136047 Rattus no
116	20	0.8	5215	6	BD136831	BD136831 Group B s	189	20	0.8	194907	2	AC129616	AC129616 Rattus no
117	20	0.8	6114	1	AB039951	AB039951 Bacillus	190	20	0.8	196107	8	ATCHRIV34	AL161534 Arabidops
118	20	0.8	6487	1	AE011427	AE011427 Leptospir	191	20	0.8	196491	9	AC092832	AC092832 Homo sapi
119	20	0.8	10579	10	AF177668	AF177668 Mus muscu	192	20	0.8	197658	2	AL646094	AL646094 Mus muscu
120	20	0.8	10618	1	AE010110	AE010110 Streptococ	193	20	0.8	197658	2	AL646094	AL646094 Mus muscu
121	20	0.8	10825	1	AE006623	AE006623 Streptococ	194	20	0.8	199050	1	SAG766850	SAG766850 Streptococ
122	20	0.8	11617	6	AX602163	AX602163 Sequence	195	20	0.8	200602	10	AL590997	AL590997 Mouse DNA
123	20	0.8	12141	1	AE014248	AE014248 Streptococ	196	20	0.8	203223	2	AC127395	AC127395 Rattus no
124	20	0.8	30799	8	SPBC19C2	SPBC19C2 Sequence	197	20	0.8	203223	2	AC127395	AC127395 Rattus no
125	20	0.8	43090	3	CRC08B6	CRC08B6 Sequence	198	20	0.8	204082	9	AC005323	AC005323 Homo sapi
126	20	0.8	50354	1	AE014169	AE014169 Streptococ	199	20	0.8	205667	2	AC118762	AC118762 Rattus no
127	20	0.8	52144	2	AC104213	AC104213 Mus muscu	200	20	0.8	206624	9	AP001825	AP001825 Homo sapi
128	20	0.8	60683	2	AL356312	AL356312 Human DNA							
129	20	0.8	61957	9	AL135922	AL135922 Human DNA							
130	20	0.8	64722	2	AC100472	AC100472 Mus muscu							
131	20	0.8	75059	9	AC125418	AC125418 Homo sapi							
132	20	0.8	75668	8	AP006097	AP006097 Lotus japi							
133	20	0.8	77405	2	AL353634	AL353634 Homo sapi							
134	20	0.8	79676	8	AB013396	AB013396 Arabidops							
135	20	0.8	91803	9	AC092318	AC092318 Homo sapi							
136	20	0.8	101981	9	AC003001	AC003001 Homo sapi							
137	20	0.8	103476	2	AC027350	AC027350 Homo sapi							
138	20	0.8	103505	9	AC026788	AC026788 Homo sapi							

ALIGNMENTS

RESULT 1
LOCUS ARI20265
DEFINITION Sequence 55 from patent US 6159469.
ACCESSION ARI20265
VERSION ARI20265.1 GI:14103841

2389 bp DNA
linear PAT 16-MAY-2001

KEYWORDS Unknown.
 SOURCE Unknown.
 ORGANISM Unclassified.
 REFERENCE 1 (bases 1 to 2389)
 AUTHORS Choi, G.H., Kunsch, C.A., Barash, S.C., Dillon, P.J., Dougherty, B.,
 Fannon, M.R. and Rosen, C.A.
 TITLE Streptococcus pneumoniae antigens and vaccines
 JOURNAL Patent: US 6159469-A 55 12-DEC-2000;
 FEATURES location/Qualifiers
 source 1..2389
 BASE COUNT 830 a 461 c 486 g 611 t 1 others
 ORIGIN

Query Match 100.0%; Score 2389; DB 6; Length 2389;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAAAGGTTAGAAATAATCGTGTCTTA
DB 1 TTCTTACGAGTTGGAGCTGTATCAAGCTAGAAAGGTTAGAAATAATCGTGTCTTA 60
QY 61 TATAGATGAAAAACAAGCAGCAAAACGAGAAATTTGACTCTGATGAGGTTAGCAA
DB 61 TATAGATGAAAAACAAGCAGCAAAACGAGAAATTTGACTCTGATGAGGTTAGCAA 120
QY 121 GGGTGAAGAAATCAATGCTGAGCAAAATCGTCATCAAGATACAGACCAAGGCTATGTCAC
DB 121 GGGTGAAGAAATCAATGCTGAGCAAAATCGTCATCAAGATACAGACCAAGGCTATGTCAC 180
QY 181 TTCACATGCGACCACTATCATTTATACAAATGTTAGGTTCCCTTATGACGCTATCATGAC
DB 181 TTCACATGCGACCACTATCATTTATACAAATGTTAGGTTCCCTTATGACGCTATCATGAC 240
QY 241 TGAAGAAATTAATCATGATAAAGATCCAAACTATAAGCTAAAAAGATGAGATATTGTTATGA
DB 241 TGAAGAAATTAATCATGATAAAGATCCAAACTATAAGCTAAAAAGATGAGATATTGTTATGA 300
QY 301 GGTCAAGGTTGATATGTTATCAAGGTAGATGAAAAATCTATGTTTACTTAAAGATGTC
DB 301 GGTCAAGGTTGATATGTTATCAAGGTAGATGAAAAATCTATGTTTACTTAAAGATGTC 360
QY 361 TGCCCAACGCGGATACGTCCTGTAACAAGGAAATCAATTCACAAAAACAAGAGCATG
DB 361 TGCCCAACGCGGATACGTCCTGTAACAAGGAAATCAATTCACAAAAACAAGAGCATG 420
QY 421 TCAACATCGTGAAGGTGAACCTCAAGAAACGATGCTGCTGCTTGCGACAGTTGCGA
DB 421 TCAACATCGTGAAGGTGAACCTCAAGAAACGATGCTGCTGCTTGCGACAGTTGCGA 480
QY 481 AGAGCGCTTACTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGATAC
DB 481 AGAGCGCTTACTACAGATGATGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
QY 541 TGGTATGCTTATATGCTTCTCATGAGATCATTAACATTAACATTCCTTAAGATGAGTT
DB 541 TGGTATGCTTATATGCTTCTCATGAGATCATTAACATTAACATTCCTTAAGATGAGTT 600
QY 601 ATCAGCTACGAGTTGGCTGCTGCGACAAACCTTCTATCTGTCGAGAAATCTGCAAA
DB 601 ATCAGCTACGAGTTGGCTGCTGCGACAAACCTTCTATCTGTCGAGAAATCTGCAAA 660
QY 661 TTCAAGAACTTATCGCGCAAAATATGCGATTAACCTTCAAGAAACAACGCGGTACTCTTC
DB 661 TTCAAGAACTTATCGCGCAAAATATGCGATTAACCTTCAAGAAACAACGCGGTACTCTTC 720
QY 721 TGTAGCAATCCAGAACTACAAATTAATAACAACCAACAGAAACACTTAACAGTCA
DB 721 TGTAGCAATCCAGAACTACAAATTAATAACAACCAACAGAAACACTTAACAGTCA 780
QY 781 AGCAAGTCAAAAGTATGATGATGCTCTTGAACAGCTCTCAAACTGCTTTGAG
DB 781 AGCAAGTCAAAAGTATGATGATGCTCTTGAACAGCTCTCAAACTGCTTTGAG 840

```

```

DB 781 AGCAAGTCAAAAGTATGATGATGCTCTTGAACAGCTCTCAAACTGCTTTGAG 840
QY 841 TCAAGACATGATGATGATGATGCTGCTTGTATGATCAGCAAAATCAGAAGTCGAC
DB 841 TCAAGACATGATGATGATGATGATGCTGCTTGTATGATCAGCAAAATCAGAAGTCGAC 900
QY 901 AGCTAGAGTGTTCAGTGCACACGAGATCATTAACATTCCTTACTTCAAAAT
DB 901 AGCTAGAGTGTTCAGTGCACACGAGATCATTAACATTCCTTACTTCAAAAT 960
QY 961 GTCTGAATTTGAGAACGATATGCTGATATATTCCTTGTATTCGTTCAACCAATG
DB 961 GTCTGAATTTGAGAACGATATGCTGATATATTCCTTGTATTCGTTCAACCAATG 1020
QY 1021 GGTACCAATTCAGAGCAGAAACCAAGTCCACACGAGTCCGGAACCTAGTCCAGG
DB 1021 GGTACCAATTCAGAGCAGAAACCAAGTCCACACGAGTCCGGAACCTAGTCCAGG 1080
QY 1081 CCCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTGTGTTAGTCAAGTGT
DB 1081 CCCGCAACCTGACCAAAATCTTAAATAGACTCAAAATCTTGTGTTAGTCAAGTGT 1140
QY 1141 ACGAAAGTTGGGAGAGATATGATTCGAGAAAAAGGCAATCTCTGTTATGCTTTGC
DB 1141 ACGAAAGTTGGGAGAGATATGATTCGAGAAAAAGGCAATCTCTGTTATGCTTTGC 1200
QY 1201 GAAGAATTTACATCTGAAATCTGTTAAATCTTGAAGCAAGTATCAAAACAAGAG
DB 1201 GAAGAATTTACATCTGAAATCTGTTAAATCTTGAAGCAAGTATCAAAACAAGAG 1260
QY 1261 TGTTCACACACTTAACTGCTAAAAAAGAAATGTTCTCTGTCGACCAAGATTTTA
DB 1261 TGTTCACACACTTAACTGCTAAAAAAGAAATGTTCTCTGTCGACCAAGATTTTA 1320
QY 1321 TGATTAAGCATATATCTGTTAACTGAGGCTCATTAACCTGTTGAAAAATTAAGGTCG
DB 1321 TGATTAAGCATATATCTGTTAACTGAGGCTCATTAACCTGTTGAAAAATTAAGGTCG 1380
QY 1381 TAAATCTGATTTCCAGGCTTAAAGCAATTAAGAGCTGATGATGATGATGATGATG
DB 1381 TAAATCTGATTTCCAGGCTTAAAGCAATTAAGAGCTGATGATGATGATGATGATG 1440
QY 1441 TAAAGAAATTTGATGATGATTTTATGAGCTTCTTGAACCAATTAACCTCAGAGG
DB 1441 TAAAGAAATTTGATGATGATTTTATGAGCTTCTTGAACCAATTAACCTCAGAGG 1500
QY 1501 ACTTGCAAAACCAATTTCTCAATTAAGTATTAAGAGCAAGATTCGATTCAT
DB 1501 ACTTGCAAAACCAATTTCTCAATTAAGTATTAAGAGCAAGATTCGATTCAT 1560
QY 1561 AGCTGATAGTATTAACAGTCAAGTGTATCTTTTATGAAACATGATATTAACATGA
DB 1561 AGCTGATAGTATTAACAGTCAAGTGTATCTTTTATGAAACATGATATTAACATGA 1620
QY 1621 TGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG
DB 1621 TGAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1680
QY 1681 CCTTCTGATTAAGAAAAAGTTGACGCTCAAGCTTATTAAGAAAAAGGATCTTAC
DB 1681 CCTTCTGATTAAGAAAAAGTTGACGCTCAAGCTTATTAAGAAAAAGGATCTTAC 1740
QY 1741 TCCATCTCAGACGAGATGTTTAAAGCAATTCGAATGAGATGAGCAGCTATTTA
DB 1741 TCCATCTCAGACGAGATGTTTAAAGCAATTCGAATGAGATGAGCAGCTATTTA 1800
QY 1801 CAATCTGTGAAGGAGAAAAAGAAATCAATTCGATTCGATTCATATATGTTAGAGA
DB 1801 CAATCTGTGAAGGAGAAAAAGAAATCAATTCGATTCGATTCATATATGTTAGAGA 1860
QY 1861 TACAGTTGAGTTAAAAAGGTAATTTGATATTCCTCATTAAGATCATTAACATTAAT
DB 1861 TACAGTTGAGTTAAAAAGGTAATTTGATATTCCTCATTAAGATCATTAACATTAAT 1920

```

QY	1921	TAAGATTCCTGGTTGGATGATCAACAATPACAACATCAACTCAATGGCTATACCTGGAGA	198
Dp	1921	TAAATTTGCTGGTTTGATGATCAACAATPACAACATCAACTCAATGGCTATACCTGGAGA	198
QY	1981	TTTGTTCGACGATTAAGTACTAGTAGAAGAACCCCTGACGACACTCCACATTCATATGA	2040
Dp	1981	TTTGTTCGACGATTAAGTACTAGTAGAAGAACCCCTGACGACACTCCACATTCATATGA	2040
QY	2041	TGGATGGGGCAATGCGACGTGACATGTGTAGGCAAGAAAGCCACATGAAAGATCCAA	2100
Dp	2041	TGGATGGGGCAATGCGACGTGACATGTGTAGGCAAGAAAGCCACATGAAAGATCCAA	2100
QY	2101	TAAGAACTTCAAACGCGATGGAAGACCAGTAGAGGAAAACCTCGCTGACCGAGAAGTCCC	2160
Dp	2101	TAAGAACTTCAAACGCGATGGAAGACCAGTAGAGGAAAACCTCGCTGACCGAGAAGTCCC	2160
QY	2161	TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAAGAGCAGAAATTTTGGCTTGGCA	2220
Dp	2161	TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAAGAGCAGAAATTTTGGCTTGGCA	2220
QY	2221	AGTAACGATTTCTAGTCTGAAAGCCAAATGCAACAGAAACTCTAGCTGGTTACGAAATTA	2280
Dp	2221	AGTAACGATTTCTAGTCTGAAAGCCAAATGCAACAGAAACTCTAGCTGGTTACGAAATTA	2280
QY	2281	TTTGACTCTTCAAAATATGATATACAAATAGTATCATGCGAAGCAGAAAAATTACTTGC	2340
Dp	2281	TTTGACTCTTCAAAATATGATATACAAATAGTATCATGCGAAGCAGAAAAATTACTTGC	2340
QY	2341	GTTGTTAAAGGAAGTATCCCTCATCTGTGTAAGTAAGAAAAAAATTAAC	2389
Dp	2341	GTTGTTAAAGGAAGTATCCCTCATCTGTGTAAGTAAGAAAAAAATTAAC	2389

[illegible]

Db	1	TTCTTAGAGTTGGGACGTGATCAAGCTAGAACGCTTAGGAAAAATATCGTTTCCCTA	60
QY	61	TATAGATGAAAAACAAGCAGCGCAAAAAACGGAGATTGACTCTGTATGAGTTAGCA	120
Db	61	TATAGATGAAAAACAAGCAGCGCAAAAAACGGAGATTGACTCTGTATGAGTTAGCA	120
QY	121	GGCGAAGGAATCAATGCTGAGCAAAATGCTACAAAGTAAACAGCAAGGCTATGTCAC	180
Db	121	GGCGAAGGAATCAATGCTGAGCAAAATGCTACAAAGTAAACAGCAAGGCTATGTCAC	180
QY	181	TTCAATGCGCACCACTATCAATTATTAACAATGATAGGTTCTTATGACGGTATCATGAC	240
Db	181	TTCAATGCGCACCACTATCAATTATTAACAATGATAGGTTCTTATGACGGTATCATGAC	240
QY	241	TGAAGATTACTCATGAAAGATCCAAACTATAGCTAAAGATGAGATATTGTTAATGA	300
Db	241	TGAAGATTACTCATGAAAGATCCAAACTATAGCTAAAGATGAGATATTGTTAATGA	300
QY	301	GGTCAAGGATGATATGTTATCAAGGTATGAGAAAAATCATATGTTTACTTAAAGATGC	360
Db	301	GGTCAAGGATGATATGTTATCAAGGTATGAGAAAAATCATATGTTTACTTAAAGATGC	360
QY	361	TGCCCAAGCGGATTAACGTCCTGATACAAAAAGAGAAATCAATGACAAAAACAAGACATAG	420
Db	361	TGCCCAAGCGGATTAACGTCCTGATACAAAAAGAGAAATCAATGACAAAAACAAGACATAG	420
QY	421	TCAAATCTGTAAGGTGGAACTCCAGAAAAGATGATGTCGTTGCTCTTGACGTTCCGA	480
Db	421	TCAAATCTGTAAGGTGGAACTCCAGAAAAGATGATGTCGTTGCTCTTGACGTTCCGA	480
QY	481	AGGACGCTATCTACAGATGATGTTATTTTAAAGCTTCTGATATCATAGAGATAC	540
Db	481	AGGACGCTATCTACAGATGATGTTATTTTAAAGCTTCTGATATCATAGAGATAC	540
QY	541	TGGATGATGCTATATCGTTCTCATGAGGATCATTAACATTACCTTAAGATGAGTT	600
Db	541	TGGATGATGCTATATCGTTCTCATGAGGATCATTAACATTACCTTAAGATGAGTT	600
QY	601	ATCAGCTAGCAGTTGCTGCTGTCAGAAAGCCTTCTATCTGTCGAGAGAAATCTGTCAA	660
Db	601	ATCAGCTAGCAGTTGCTGCTGTCAGAAAGCCTTCTATCTGTCGAGAGAAATCTGTCAA	660
QY	661	TTCAAAGAACTATGCGCGACAAAATAGCGATTAACATTCAAGAACAACTGGTATCTTC	720
Db	661	TTCAAAGAACTATGCGCGACAAAATAGCGATTAACATTCAAGAACAACTGGTATCTTC	720
QY	721	TGTAGCAATCCAGGAATCAAAATTAACAACAAGGAAACAAGCAACAATACTAACATGCA	780
Db	721	TGTAGCAATCCAGGAATCAAAATTAACAACAAGGAAACAAGCAACAATACTAACATGCA	780
QY	781	AGCAAGTCAAAAGTAAATGACATTTGATAGTCTCTTGAAAACGCTCTTCAAACTGCTTTGAG	840
Db	781	AGCAAGTCAAAAGTAAATGACATTTGATAGTCTCTTGAAAACGCTCTTCAAACTGCTTTGAG	840
QY	841	TCAAAGCATGTAGATCTGATGCGCTTGTCTTTGATCCAGACAAATACAAAGTCGAAC	900
Db	841	TCAAAGCATGTAGATCTGATGCGCTTGTCTTTGATCCAGACAAATACAAAGTCGAAC	900
QY	901	AGCTAAGAGTGTGCAATGTCACACGAGATCATTTACACTTAATCCCTTAATCTTCAAAAT	960
Db	901	AGCTAAGAGTGTGCAATGTCACACGAGATCATTTACACTTAATCCCTTAATCTTCAAAAT	960
QY	961	GTCGAATTGGAAAGAAAGAAATCGCTGCTATATATCCCTTCTGTTATGCTTCAACCATTC	1020
Db	961	GTCGAATTGGAAAGAAAGAAATCGCTGCTATATATCCCTTCTGTTATGCTTCAACCATTC	1020
QY	1021	GGTACAGATTTCAAGGCTAGAACCAACCAAGTCCAAACGATCTCGGAACTAGTCCAGG	1080
Db	1021	GGTACAGATTTCAAGGCTAGAACCAACCAAGTCCAAACGATCTCGGAACTAGTCCAGG	1080
QY	1081	CCCGCAACTGCAACCAATCTTAAATAGACTCAAAATTTCTTCTTGTTAGTCAAGCTGCT	1140

```

Db      1081 CCCGCACTGCAACCAATCTTAAAAATGACTCAAAATCTCTTGGTTAGTCACTGCT 1140
Qy      1141 ACGAAAGTTGGGGAAGATATGTATTCGAGAAAAAGGCGATCTCTGTTATGTCTTTC 1200
Db      1141 ACGAAAGTTGGGGAAGATATGTATTCGAGAAAAAGGCGATCTCTGTTATGTCTTTC 1200
Qy      1201 GAAAGATTACATCTGAAATCTGTTAAAAATCTTGAAGCAAGTTATCAAAACAAGAG 1260
Db      1201 GAAAGATTACATCTGAAATCTGTTAAAAATCTTGAAGCAAGTTATCAAAACAAGAG 1260
Qy      1261 TGTTCACACACTTAACTGCTAAAAAAGAAATGTGCTCTGCTGACCAAGATTTTA 1320
Db      1261 TGTTCACACACTTAACTGCTAAAAAAGAAATGTGCTCTGCTGACCAAGATTTTA 1320
Qy      1321 TGATAAGCATTAATCTGTCTTAAGTGAAGGCTCATAAAGCCTTGTGNAATTAAGGTC 1380
Db      1321 TGATAAGCATTAATCTGTCTTAAGTGAAGGCTCATAAAGCCTTGTGNAATTAAGGTC 1380
Qy      1381 TAAATCTGATTTTCAAGCCTTAGACAAATTTAGAACGCTTGAATGATGAATGACTAA 1440
Db      1381 TAAATCTGATTTTCAAGCCTTAGACAAATTTAGAACGCTTGAATGATGAATGACTAA 1440
Qy      1441 TAAAGAAAAATGTGATGATTTATGGCATTTCTAGACCAATTTACCATCCAGAGC 1500
Db      1441 TAAAGAAAAATGTGATGATTTATGGCATTTCTAGACCAATTTACCATCCAGAGC 1500
Qy      1501 ACTTGGCAACCAATTTCTCAATTTAGATGATTAAGTGAAGCAAGTTGCTGATTCAT 1560
Db      1501 ACTTGGCAACCAATTTCTCAATTTAGATGATTAAGTGAAGCAAGTTGCTGATTCAT 1560
Qy      1561 AGCTGATAAGTATACACGCTCAGATGCTTACATTTTGTGTAACATGATATATCAGTGA 1620
Db      1561 AGCTGATAAGTATACACGCTCAGATGCTTACATTTTGTGTAACATGATATATCAGTGA 1620
Qy      1621 TGAAGAGATGATATGTAACGCTCATATGAGCCATATGCTGATTTGAAAAAGATG 1680
Db      1621 TGAAGAGATGATATGTAACGCTCATATGAGCCATATGCTGATTTGAAAAAGATG 1680
Qy      1681 CTTTCTGATAGAAAAAGTGAAGCTCAAGCTCATATTAAGAAAAAGGATCTCTACC 1740
Db      1681 CTTTCTGATAGAAAAAGTGAAGCTCAAGCTCATATTAAGAAAAAGGATCTCTACC 1740
Qy      1741 TCCATCTCCAGACGAGATGTTAAAGCAATCAACTGAGATAGTGCAGCAGCTATTTA 1800
Db      1741 TCCATCTCCAGACGAGATGTTAAAGCAATCAACTGAGATAGTGCAGCAGCTATTTA 1800
Qy      1801 CAATGCTGTGAAGGGGAAAAAGCAATTCACCTGCTGACATTCATATATGTTGAGCA 1860
Db      1801 CAATGCTGTGAAGGGGAAAAAGCAATTCACCTGCTGACATTCATATATGTTGAGCA 1860
Qy      1861 TACAGTTGAGTTAAAAAGGTAATTTGATTTCTCTCAATAGGATCATTAATATAT 1920
Db      1861 TACAGTTGAGTTAAAAAGGTAATTTGATTTCTCTCAATAGGATCATTAATATAT 1920
Qy      1921 TAAATTTGCTGTTGATGATCAACATACAAAGCTCCAAATGCTATACCTTGAAGA 1980
Db      1921 TAAATTTGCTGTTGATGATCAACATACAAAGCTCCAAATGCTATACCTTGAAGA 1980
Qy      1981 TTTGTTTGGCAGATTAAGTACTAGTGAGAACACCTGACGAACCTCCACATTTCAATGA 2040
Db      1981 TTTGTTTGGCAGATTAAGTACTAGTGAGAACACCTGACGAACCTCCACATTTCAATGA 2040
Qy      2041 TGGATGGGGCAATGCGAGTGAACAGTGTGTAAGCAAGAAAGCAACAGTGAAGATCCAA 2100
Db      2041 TGGATGGGGCAATGCGAGTGAACAGTGTGTAAGCAAGAAAGCAACAGTGAAGATCCAA 2100
Qy      2101 TAAAGACTTCAAGCGAGTGAAGAGCAGTGAAGAAACACCTGCTGAGCCAGAACTCC 2160
Db      2101 TAAAGACTTCAAGCGAGTGAAGAGCAGTGAAGAAACACCTGCTGAGCCAGAACTCC 2160
Qy      2161 TCAAGTAGAGCTGAAAAAGTGAAGGCCCACTCAAGAGCAAGAGTTTCTTGGCA 2220
Db      2161 TCAAGTAGAGCTGAAAAAGTGAAGGCCCACTCAAGAGCAAGAGTTTCTTGGCA 2220

```

```

Qy      2221 AGTAACGATTTCTAGTCTGAAGGCAATGCAACAGAAACTAGCTGCTTACGAAATTA 2280
Db      2221 AGTAACGATTTCTAGTCTGAAGGCAATGCAACAGAAACTAGCTGCTTACGAAATTA 2280
Qy      2281 TTTGACTCTCAATTTATGATGTAACATAGTATGATGCGAGAGCAAGAAAAATTA 2340
Db      2281 TTTGACTCTCAATTTATGATGTAACATAGTATGATGCGAGAGCAAGAAAAATTA 2340
Qy      2341 GTTGTTAAAGAGTATTCCTCATCTGTAAGTGAAGAAAAATTAAC 2389
Db      2341 GTTGTTAAAGAGTATTCCTCATCTGTAAGTGAAGAAAAATTAAC 2389

RESULT 3
AF291695
LOCUS      AF291695                2541 bp    DNA          linear    BCT 15-MAR-2001
DEFINITION Streptococcus pneumoniae pneumococcal histidine triad A protein
            (phtA) gene, complete cds.
ACCESSION  AF291695
VERSION    AF291695.1  GI:13345012
KEYWORDS   Streptococcus pneumoniae
SOURCE      Streptococcus pneumoniae
            Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
            Streptococcus.
REFERENCE   1 (bases 1 to 2541)
            Witzmann,T.M., Heinrichs,J.H., Adamou,J.E., Erwin,A.L., Kunsch,C.,
            Choi,G.H., Barash,S.C., Rosen,C.A., Masure,H.R., Tuomanen,E.,
            Gayle,A., Brewin,Y.A., Walsh,W., Barron,P., Lathigra,R., Hanson,M.,
            Langermann,S., Johnson,S. and Koenig,S.
            Use of a whole genome approach to identify vaccine molecules
            affecting protection against Streptococcus pneumoniae infection
            Infect. Immun. 69 (3), 1593-1598 (2001)

JOURNAL    MEDLINE
            21116976
PUBMED     11179332
REFERENCE   2 (bases 1 to 2541)
            Chol,G.H.
            Direct Submission
            Submitted (01-AUG-2000) Molecular Biology, Human Genome Sciences,
            Inc., 9410 Key West Ave., Rockville, MD 20850, USA
            Location/Qualifiers
                1..2541
                /organism="Streptococcus pneumoniae"
                /mol_type="genomic DNA"
                /strain="M4"
                /db_xref="taxon:1313"
                91..2541
                /gene="phtA"
                91..2541
                /gene="phtA"
                /note="phtA"
                /codon_start=1
                /transl_table=1
                /product="pneumococcal histidine triad A protein"
                /protein_id="AAK19155.1"
                /db_xref="GI:13345013"

FEATURES             source
     gene
     CDS

```

BASE COUNT

888 a 476 c 516 g 660 t 1 others

ORIGIN

Query Match 100.0%; Score 2389; DB 1; Length 2541;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy 1 TTCTTACGAGTGGAGCTGTATCAAGCTAGAAACGGTTAAGAAATAATCGTGTTCCTA 60
Db 150 TTCTTACGAGTGGAGCTGTATCAAGCTAGAAACGGTTAAGAAATAATCGTGTTCCTA 209
Qy 61 TATAGATGAAAAACAAGGACCCAAAAACGAGAAATTTGACTCCTGTATGAGGTTAGCA 120
Db 210 TATAGATGAAAAACAAGGACCCAAAAACGAGAAATTTGACTCCTGTATGAGGTTAGCA 269
Qy 121 GCGTGAAGGAATCAATGCTGAGCAAAATGCTCATCAAGATAACAGACCAAGGCTATGTAC 180
Db 270 GCGTGAAGGAATCAATGCTGAGCAAAATGCTCATCAAGATAACAGACCAAGGCTATGTAC 329
Qy 181 TTCAATGAGGACCACTATCATTTATTAAGATGTAAGGTTCCCTTATGACGCTATCATCAG 240
Db 330 TTCAATGAGGACCACTATCATTTATTAAGATGTAAGGTTCCCTTATGACGCTATCATCAG 389
Qy 241 TGAAGAAATTAATCATGAAAGATCCAACTATAGCTTAAAGATGAGATATTGTTAATGA 300
Db 390 TGAAGAAATTAATCATGAAAGATCCAACTATAGCTTAAAGATGAGATATTGTTAATGA 449
Qy 301 GGTCAAGGATGATATGTTATCAAGTATGATGAATAATCTATGTTTACCTTAAGAGATGC 360
Db 450 GGTCAAGGATGATATGTTATCAAGTATGATGAATAATCTATGTTTACCTTAAGAGATGC 509
Qy 361 TGCCCAACGCGATTAACGTCCTGTAACAAAAGAGAAATCAATGCACAAAAACAAGACATAG 420
Db 510 TGCCCAACGCGATTAACGTCCTGTAACAAAAGAGAAATCAATGCACAAAAACAAGACATAG 569
Qy 421 TCAACATGTTGAAGGTGGAATCTCAAGAAACGATGTGTCTGTGCTTGGCCTTGGCA 480
Db 570 TCAACATGTTGAAGGTGGAATCTCAAGAAACGATGTGTCTGTGCTTGGCCTTGGCA 629
Qy 481 AGGAGCTATATCAAGATGATGTTATATCTTTAATGCTTGTATCATAGAGATAG 540
Db 630 AGGAGCTATATCAAGATGATGTTATATCTTTAATGCTTGTATCATAGAGATAG 689
Qy 541 TGGTATGCTTATATCGTTCCCTCATGAGATCATTAACATTAATCTCTTAAGATGATG 600
Db 690 TGGTATGCTTATATCGTTCCCTCATGAGATCATTAACATTAATCTCTTAAGATGATG 749
Qy 601 ATCACTAGCGAGTGGCTGCTGCAAGACCTTCTTATCTGTGTGAGAGAAATCTGTCAA 660
Db 750 ATCACTAGCGAGTGGCTGCTGCAAGACCTTCTTATCTGTGTGAGAGAAATCTGTCAA 809
Qy 661 TTCAGAACTATGCGCGCAAAATAGGGAATTAACCTTCAAGAACAACTGGGTACCTTC 720
Db 810 TTCAGAACTATGCGCGCAAAATAGGGAATTAACCTTCAAGAACAACTGGGTACCTTC 869
Qy 721 TGTAAAGCAATCCAGAACTACAATACTAACAACAAGAACAACTAACAAGTCA 780
Db 870 TGTAAAGCAATCCAGAACTACAATACTAACAACAAGAACAACTAACAAGTCA 929
Qy 781 AGCAAGTCAAGTATGACATGATGATGCTTGAAGACGCTTCAAAATGCTCTTGAAG 840
Db 930 AGCAAGTCAAGTATGACATGATGATGCTTGAAGACGCTTCAAAATGCTCTTGAAG 899
Qy 841 TCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
Db 990 TCAACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1049
Qy 901 AGCTAGAGTGTGAGTGCACACAGGAGATCATTAACATTCCTTACTCTCAAT 960
Db 1050 AGCTAGAGTGTGAGTGCACACAGGAGATCATTAACATTCCTTACTCTCAAT 1109
Qy 961 GTCTGAATTTGGAAGAAAGATGCTGTATTAATTCCTTGGTTATGCTTCAACCATG 1020
Db 1110 GTCTGAATTTGGAAGAAAGATGCTGTATTAATTCCTTGGTTATGCTTCAACCATG 1169
```

```
Qy 1021 GGTACAGATTTCAAGGCCAGAACACCAAGTCCAAACCGACTCCGGAACCTAGTCCAGG 1080
Db 1170 GGTACAGATTTCAAGGCCAGAACACCAAGTCCAAACCGACTCCGGAACCTAGTCCAGG 1229
Qy 1081 CCGCAACCTGCACCAATCTTTAAATATAGACTCAATTTCTTCTGTTAGTCACTGCT 1140
Db 1230 CCGCAACCTGCACCAATCTTTAAATATAGACTCAATTTCTTCTGTTAGTCACTGCT 1289
Qy 1141 ACGAAAGTTGGGGAAGATATGATTCGAAGAAAGGGGATCTCTGTTATGCTTCTG 1200
Db 1290 ACGAAAGTTGGGGAAGATATGATTCGAAGAAAGGGGATCTCTGTTATGCTTCTG 1349
Qy 1201 GAAAGATTTTCACTGTAACCTGTTAAATCTTTGAAAGCAAGTTATCAAAACAAGAG 1260
Db 1350 GAAAGATTTTCACTGTAACCTGTTAAATCTTTGAAAGCAAGTTATCAAAACAAGAG 1409
Qy 1261 TGTTCACACCTTTAATCTGTAACCAATATTTAGAAAGGATGCTCTCTGTAACCAAGATTTTA 1320
Db 1410 TGTTCACACCTTTAATCTGTAACCAATATTTAGAAAGGATGCTCTCTGTAACCAAGATTTTA 1469
Qy 1321 TGAATTAAGCATATTAATGTTAATCTGAGGCTCAATTAAGCCTGTTGNAATTAAGGCTG 1380
Db 1470 TGAATTAAGCATATTAATGTTAATCTGAGGCTCAATTAAGCCTGTTGNAATTAAGGCTG 1529
Qy 1381 TAAATTCGATTTTCAAGCCTTACACAAATTTATGAACGCTTGAAATGATGATGACTAA 1440
Db 1530 TAAATTCGATTTTCAAGCCTTACACAAATTTATGAACGCTTGAAATGATGATGACTAA 1589
Qy 1441 TAAAGAAATTTGGTATGATGATTTATTTGGCAATTCCTAGACCAATTCACAGAGCG 1500
Db 1590 TAAAGAAATTTGGTATGATGATTTATTTGGCAATTCCTAGACCAATTCACAGAGCG 1649
Qy 1501 ACTTGGAAACCAATTCCTCAATTTGATGATACGAAGAGAGAGTGTATGCTCAAT 1560
Db 1650 ACTTGGAAACCAATTCCTCAATTTGATGATACGAAGAGAGAGTGTATGCTCAAT 1709
Qy 1561 AGCTGATTAAGTATCAACGTCAGATGTTACATTTTGTATGAACATGATTAATCACTGA 1620
Db 1710 AGCTGATTAAGTATCAACGTCAGATGTTACATTTTGTATGAACATGATTAATCACTGA 1769
Qy 1621 TGAAGAGATGATATGTAACGCTCATATATGGGCAATGATGATGATGATGATGATG 1680
Db 1770 TGAAGAGATGATATGTAACGCTCATATATGGGCAATGATGATGATGATGATGATG 1829
Qy 1681 CCTTTCGATTAAGAAACGTTGAGCTCAAGCCTTATATTAAGAAACGTTATCTTACC 1740
Db 1830 CCTTTCGATTAAGAAACGTTGAGCTCAAGCCTTATATTAAGAAACGTTATCTTACC 1889
Qy 1741 TCCATCTCCAGACGATGTTAAAGCAATTCCAATCGAGATATGTCACAGCTATTTTA 1800
Db 1890 TCCATCTCCAGACGATGTTAAAGCAATTCCAATCGAGATATGTCACAGCTATTTTA 1949
Qy 1801 CAATGCTGTGAAGGGGAAAAACGAATTCATCTGTTGCACTTCCATATATGTTGAGCA 1860
Db 1950 CAATGCTGTGAAGGGGAAAAACGAATTCATCTGTTGCACTTCCATATATGTTGAGCA 2009
Qy 1861 TACAGTTAGGTTAAACCGTAATTTGATTAATTCCTCATTAAGGATATTAACATATAT 1920
Db 2010 TACAGTTAGGTTAAACCGTAATTTGATTAATTCCTCATTAAGGATATTAACATATAT 2069
Qy 1921 TAAATTTGCTTGGTTTATGATCACAATCAAAAGCTCCAAATGGCTATACCTTGAAGA 1980
Db 2070 TAAATTTGCTTGGTTTATGATCACAATCAAAAGCTCCAAATGGCTATACCTTGAAGA 2129
Qy 1981 TTTGTTTGCAGATTAAGTATCTACGTAGAACCTGACGAAGCTGCACATTTCTAATGA 2040
Db 2130 TTTGTTTGCAGATTAAGTATCTACGTAGAACCTGACGAAGCTGCACATTTCTAATGA 2189
Qy 2041 TGAATGGGGCAATGCTGAGCATGTTGTTAGGCAAGAAAGACCAACGTAAGATCCAAA 2100
Db 2190 TGAATGGGGCAATGCTGAGCATGTTGTTAGGCAAGAAAGACCAACGTAAGATCCAAA 2249
```

QY 2101 TAAAGATTCAAGCGGATGAAGAGCCAGTAGAGAAACCTGCTGACCGAAGACTCC 2160
 DB 2250 TAAAGACTTCAAGCGGATGAAGAGCCAGTAGAGAAACCTGCTGACCGAAGACTCC 2309
 QY 2161 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGAAAGTTTGTGTCGGA 2220
 DB 2310 TCAAGTAGAGACTGAAAAAGTAGAAGCCCACTCAAGAGAAAGTTTGTGTCGGA 2369
 QY 2221 AGTAACGAGATTCTAGCTGAAAAAGCCAACTGCAAGAACTCTAGCTGTTTACGAATA 2280
 DB 2370 AGTAACGAGATTCTAGCTGAAAAAGCCAACTGCAAGAACTCTAGCTGTTTACGAATA 2429
 QY 2281 TTTGACTCTTCAATTAATGATTAACAATAGTATCATGCGAAGAGAAAAATTACTTGC 2340
 DB 2430 TTTGACTCTTCAATTAATGATTAACAATAGTATCATGCGAAGAGAAAAATTACTTGC 2489
 QY 2341 GTTGTAAAAAGAGTAATCTTCTCATCTGTAGTAAGAAAAATAATAC 2389
 DB 2490 GTTGTAAAAAGAGTAATCTTCTCATCTGTAGTAAGAAAAATAATAC 2538

RESULT 4
 AX569139 2406 bp DNA linear PAT 29-NOV-2002

LOCUS AX569139 Sequence 2347 from Patent WO02077021.
 DEFINITION AX569139
 ACCESSION AX569139
 VERSION AX569139.1 GI:26002636
 KEYWORDS

SOURCE Streptococcus pneumoniae
 ORGANISM Streptococcus pneumoniae
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.

REFERENCE 1
 AUTHORS Masignani,V., Tettelin,H. and Fraser,C.
 TITLE Streptococcus pneumoniae proteins and nucleic acids
 JOURNAL Patent: WO 02077021-A 2347 03-OCT-2002;
 Chiron Spa (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
 FEATURES
 source 1..2406
 Location/Qualifiers
 /organism="Streptococcus pneumoniae"
 /mol_type="genomic DNA"
 /db_xref="taxon:1313"

BASE COUNT 834 a 461 c 490 g 621 t
 ORIGIN

Query Match 97.9%; Score 2338; DB 6; Length 2406;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTCTTCAGAGTTGGAGCTGATCAAGCTGAACGGTTAAGAAATTAATCGTTTCTTA 60
 DB 18 TTCTTCAGAGTTGGAGCTGATCAAGCTGAACGGTTAAGAAATTAATCGTTTCTTA 77
 QY 61 TATAGATGAAAAACAAGCAGCGCAAAAAACGAGAAATTTGACTCTGATGAGGTTAGCA 120
 DB 78 TATAGATGAAAAACAAGCAGCGCAAAAAACGAGAAATTTGACTCTGATGAGGTTAGCA 137
 QY 121 GCGTGAAGAAATCAATGCTGAGCAATTCGTCAATCAAGATPACAGCAAGGCTATGTC 180
 DB 138 GCGTGAAGAAATCAATGCTGAGCAATTCGTCAATCAAGATPACAGCAAGGCTATGTC 197
 QY 181 TTTCACATGGCGACCACTATCATTTATTAACAATGTTAGGTTCCCTTATGAGGCTATCAT 240
 DB 198 TTTCACATGGCGACCACTATCATTTATTAACAATGTTAGGTTCCCTTATGAGGCTATCAT 257
 QY 241 TGAAGATTAATCAATGAAGATCAAACTATAAGCTTAAGATGAAGATATTGTTATGA 300
 DB 258 TGAAGATTAATCAATGAAGATCAAACTATAAGCTTAAGATGAAGATATTGTTATGA 317
 QY 301 GGTCAAGGTTGATGTTATCAAGGTAAGTGAAGAAATTAATGTTTACCTTAAGATGC 360
 DB 318 GGTCAAGGTTGATGTTATCAAGGTAAGTGAAGAAATTAATGTTTACCTTAAGATGC 377

QY 361 TGCCACCGCGATTAACGTCCGTAACAAAAAGAGAAATCAATGCACAAAAACAAGACATAG 420
 DB 378 TGCCACCGCGATTAACGTCCGTAACAAAAAGAGAAATCAATGCACAAAAACAAGACATAG 437
 QY 421 TCAATCATGTGAAGGTGGAATCTCAAGAAACGATGTGTGTGCTTGGCAGTTCCGA 480
 DB 438 TCAATCATGTGAAGGTGGAATCTCAAGAAACGATGTGTGTGCTTGGCAGTTCCGA 497
 QY 481 AGGAGCTATPCTAAGATGATGTTATATCTTTAATGCTTGTGATATCATAGAGATAC 540
 DB 498 AGGAGCTATPCTAAGATGATGTTATATCTTTAATGCTTGTGATATCATAGAGATAC 557
 QY 541 TGTGATGCTTATATGTTTCTCATGAGATCATTAACATTCTCTTAAGAAAGATT 600
 DB 558 TGTGATGCTTATATGTTTCTCATGAGATCATTAACATTCTCTTAAGAAAGATT 617
 QY 601 ATCAGCTAGCGAGTTGCTGCTGCAAGACCTTCTTATCTGTGTCAGAGAAATCTGTCAA 660
 DB 618 ATCAGCTAGCGAGTTGCTGCTGCAAGACCTTCTTATCTGTGTCAGAGAAATCTGTCAA 677
 QY 661 TTCAGAACTATGCGCGCAAAAAATAGGATTAACCTTCAAGAACAACTGGGTACCTTC 720
 DB 678 TTCAGAACTATGCGCGCAAAAAATAGGATTAACCTTCAAGAACAACTGGGTACCTTC 737
 QY 721 TGTAGCAATCCAGAACTACAAATTAACAACAAGCAACAACAGCAACATCAAGATCA 780
 DB 738 TGTAGCAATCCAGAACTACAAATTAACAACAAGCAACAACAGCAACATCAAGATCA 797
 QY 781 AGCAAGTCAAGTAATGACATTGATAGTCTTTGAAACAGCTTCAAACTGCTTTGAG 840
 DB 798 AGCAAGTCAAGTAATGACATTGATAGTCTTTGAAACAGCTTCAAACTGCTTTGAG 857
 QY 841 TCAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900
 DB 858 TCAAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 917
 QY 901 AGCTAGAGTGTGTCAGTGCACACAGAGATCAATACATTCCTTACTCTCAAT 960
 DB 918 AGCTAGAGTGTGTCAGTGCACACAGAGATCAATACATTCCTTACTCTCAAT 977
 QY 961 GTCTGAATGGAAGACGAATGCTGATTAATTCCTTGTGTTATGCTTCAACCATTTG 1020
 DB 978 GTCTGAATGGAAGACGAATGCTGATTAATTCCTTGTGTTATGCTTCAACCATTTG 1037
 QY 1021 GGTACCAATTCAGAGCCAGAAACAACAGTCCACACAGCTCCGGAACCTAGTCCAG 1080
 DB 1038 GGTACCAATTCAGAGCCAGAAACAACAGTCCACACAGCTCCGGAACCTAGTCCAG 1097
 QY 1081 CCCGCACTGCAACCAATCTTAATATGACTCAAAATCTTCTTGGTTAGTACGCTGGT 1140
 DB 1098 CCCGCACTGCAACCAATCTTAATATGACTCAAAATCTTCTTGGTTAGTACGCTGGT 1157
 QY 1141 ACGAAAGTTGGGAGAGATGATGTAATCGAAGAAAGGCACTCTGTTATGTCTTTGC 1200
 DB 1158 ACGAAAGTTGGGAGAGATGATGTAATCGAAGAAAGGCACTCTGTTATGTCTTTGC 1217
 QY 1201 GAAAGATTTACATCTGAAACGTGTTAAATTTGAAAGCAAGTTATCAAAACAAGAG 1260
 DB 1218 GAAAGATTTACATCTGAAACGTGTTAAATTTGAAAGCAAGTTATCAAAACAAGAG 1277
 QY 1261 TGTTCACACACTTAACTGCTTAAGAAAGAAATGTTGCTCTGCGAACAAGATTTTA 1320
 DB 1278 TGTTCACACACTTAACTGCTTAAGAAAGAAATGTTGCTCTGCGAACAAGATTTTA 1337
 QY 1321 TGAATAAGCATTAATCTGTTAACTGAGGCTCATTAAGGCTTGTGTAATTAAGGTTG 1380
 DB 1338 TGAATAAGCATTAATCTGTTAACTGAGGCTCATTAAGGCTTGTGTAATTAAGGTTG 1397
 QY 1381 TAAATCTGATTTCCAGGCTTGAACAATTAATTAAGGCTGAATGATGATGATGATGAT 1440
 DB 1398 TAAATCTGATTTCCAGGCTTGAACAATTAATTAAGGCTGAATGATGATGATGATGAT 1457
 QY 1441 TAAAGAAAAATGATGATGATTTATGTCATTCCTAGACCAATTAACCATCAAGAGC 1500

```

Db      1458 TAAAGAAAAATGGTAGATGATTTATGTCATCTCTAGACCAATTAACCAATCCAGACG 1517
Oy      1501 ACTTGCAAAACCAATTTCTCAATGAGATAGTGAAGCAAGAGTTCGATGCTCAAT 1560
Db      1518 ACTTGCAAAACCAATTTCTCAATGAGATAGTGAAGCAAGAGTTCGATGCTCAAT 1577
Oy      1561 AGCTGATTAAGTATACACGTCAGATGTTACATTTTGTGAACATGATATATACATGA 1620
Db      1578 AGCTGATTAAGTATACACGTCAGATGTTACATTTTGTGAACATGATATATACATGA 1637
Oy      1621 TGAAGAGATGATATATGTAACCCCTCATTTGGGCCATAGTCACTGATTTGAAAAATAG 1680
Db      1638 TGAAGAGATGATATATGTAACCCCTCATTTGGGCCATAGTCACTGATTTGAAAAATAG 1697
Oy      1681 CTTTCTGTAGTAAAGAAAAATGGACGCTCAAGCTATATCTAAAGAAAAAGTATCTACC 1740
Db      1698 CTTTCTGTAGTAAAGAAAAATGGACGCTCAAGCTATATCTAAAGAAAAAGTATCTACC 1757
Oy      1741 TCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGAGATAGTGCAGACGATTTTA 1800
Db      1758 TCCATCTCCAGACGAGATGTTAAAGCAAAATCCAACTGAGATAGTGCAGACGATTTTA 1817
Oy      1801 CAATGCTGTAAGAGGGGAAAAAGCAATTCACCTGTTGACCTTCCATATATGTTGAGCA 1860
Db      1818 CAATGCTGTAAGAGGGGAAAAAGCAATTCACCTGTTGACCTTCCATATATGTTGAGCA 1877
Oy      1861 TACAGTTGAGGTTAAACCGGTAATTTGATTTATCTCTCATAGATTCATATCATATAT 1920
Db      1878 TACAGTTGAGGTTAAACCGGTAATTTGATTTATCTCTCATAGATTCATATCATATAT 1937
Oy      1921 TAAATTTGCTGTTGTTGATGATCACAACATACAAAGCTCCAAATGCTATACCTTGAAAG 1980
Db      1938 TAAATTTGCTGTTGTTGATGATCACAACATACAAAGCTCCAAATGCTATACCTTGAAAG 1997
Oy      1981 TTTGTTTGCAGAGTTAAGTATCTAGTAGAACAACCTGACGAAGCTTCCATTTCAATGA 2040
Db      1998 TTTGTTTGCAGAGTTAAGTATCTAGTAGAACAACCTGACGAAGCTTCCATTTCAATGA 2057
Oy      2041 TGGATGGGGCAATGCGAGTAGAGATGTTTGGGCAAGAAAGCAACAGTAAAGATCCAAA 2100
Db      2058 TGGATGGGGCAATGCGAGTAGAGATGTTTGGGCAAGAAAGCAACAGTAAAGATCCAAA 2117
Oy      2101 TAAAGACTTCAAAAGCGGATGAAGAGCCAGTAGAGAAAAACCTGCTGAGCCAGAGTCCC 2160
Db      2118 TAAAGACTTCAAAAGCGGATGAAGAGCCAGTAGAGAAAAACCTGCTGAGCCAGAGTCCC 2177
Oy      2161 TCAAGTAGAGACTGAAAAAGTGAAGAGCCCACTCAAGAGAGCAAGATTGTTGCTTGCAG 2220
Db      2178 TCAAGTAGAGACTGAAAAAGTGAAGAGCCCACTCAAGAGAGCAAGATTGTTGCTTGCAG 2237
Oy      2221 AGTAAAGGATCTACTCTGAAAGGCAATGCAACAAAACCTCTAGTGGTTTACGAAATTA 2280
Db      2238 AGTAAAGGATCTACTCTGAAAGGCAATGCAACAAAACCTCTAGTGGTTTACGAAATTA 2297
Oy      2281 TTTGACTCTTCAATATGATGATACATATGATCTAGTGCAGAGAGCAAAAAATTTACTGC 2340
Db      2298 TTTGACTCTTCAATATGATGATACATATGATCTAGTGCAGAGAGCAAAAAATTTACTGC 2357
Oy      2341 GTTGTAAAGAGAGTAATCTTCTCATCTGTAAAGTAAAGAAAAATTAAC 2389
Db      2358 GTTGTAAAGAGAGTAATCTTCTCATCTGTAAAGTAAAGAAAAATTAAC 2406

```

```

RESULT 5
LOCUS      AR218862      8195 bp      DNA      linear      PAT 25-SEP-2002
DEFINITION Sequence 94 from patent US 6420135.
ACCESSION  AR218862
VERSION     AR218862.1  GI:23319796
KEYWORDS
SOURCE      Unknown.
ORGANISM    Unknown.

```

```

REFERENCE 1 (bases 1 to 8195)
AUTHORS   Kunsch,C.A., Choi,G.H., Dillon,P.S., Rosen,C.A., Barash,S.C.,
           Fannon,M.R. and Dougherty,B.A.
TITLE      Streptococcus pneumoniae polynucleotides and sequences
JOURNAL    Patent: US 6420135-A 94 16-JUL-2002;
FEATURES   Location/Qualifiers
            source
            1..8195
            /organism="unknown"
BASE COUNT 2688 a 1622 c 1777 g 2105 t      3 others
ORIGIN
Query Match      97.9%; Score 2338; DB 6; Length 8195;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy      1 TTTCTTACAGTTGGGACCTGTATCAAGCTAGAACGGTTAAGGAAATTAATCGTGTTCCTA 60
Db      3053 TTCTTACAGTTGGGACCTGTATCAAGCTAGAACGGTTAAGGAAATTAATCGTGTTCCTA 3112
Oy      61 TATAGATGAAAAAACAAGGACGCAAAAAACGAGAAATTTGACTCTGATGAGTTAGCA 120
Db      3113 TATAGATGAAAAAACAAGGACGCAAAAAACGAGAAATTTGACTCTGATGAGTTAGCA 3172
Oy      121 GCGTGAAGAAATCAATGCTGAGCAAAATGCTATCAAGATTAACAGACCAAGGCTATGTCAC 180
Db      3173 GCGTGAAGAAATCAATGCTGAGCAAAATGCTATCAAGATTAACAGACCAAGGCTATGTCAC 3232
Oy      181 TTCAATGCGGACCAATATCTTTTCAATAGTATAGTTCCTATACCGTATATCATAG 240
Db      3233 TTCAATGCGGACCAATATCTTTTCAATAGTATAGTTCCTATACCGTATATCATAG 3292
Oy      241 TGAAGAAATTTACTCATGAAGAATCCAAACTATTAAGCTAAAGATAGATATTTGTTAA 300
Db      3293 TGAAGAAATTTACTCATGAAGAATCCAAACTATTAAGCTAAAGATAGATATTTGTTAA 3352
Oy      301 GGTCAAGGGTGATATGTTATCAAGTAGATGAAAAATCTATGTTTACTTAAGATGC 360
Db      3353 GGTCAAGGGTGATATGTTATCAAGTAGATGAAAAATCTATGTTTACTTAAGATGC 3412
Oy      361 TGCCCAAGCGGATTAACGTCCCTACAAAAGAGAAATCAATGCAAAAAACAAGGCAATAG 420
Db      3413 TGCCCAAGCGGATTAACGTCCCTACAAAAGAGAAATCAATGCAAAAAACAAGGCAATAG 3472
Oy      421 TCAACATCTGTAAGGTGAACTCCCAAGAAACGATGCTGTTGCTTGACGACGTTGCA 480
Db      3473 TCAACATCTGTAAGGTGAACTCCCAAGAAACGATGCTGTTGCTTGACGACGTTGCA 3532
Oy      481 AGAAGCTATATCTACAGATAGTGTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
Db      3533 AGAAGCTATATCTACAGATAGTGTATATCTTTAATGCTTCTGATATCATAGAGATAC 3592
Oy      541 TGTGATGCTTATATCGTTCCTCATGAGATCATTTACCATTAACCTTCTAAGATGATT 600
Db      3593 TGTGATGCTTATATCGTTCCTCATGAGATCATTTACCATTAACCTTCTAAGATGATT 3652
Oy      601 ATCAGCTAGAGAGTTGGCTGCTGAGAGAGCTTCTCTATCTGCTGAGAGAAATCTGTCAA 660
Db      3653 ATCAGCTAGAGAGTTGGCTGCTGAGAGAGCTTCTCTATCTGCTGAGAGAAATCTGTCAA 3712
Oy      661 TTCAAGAACCTATGCGCGACAAAATAGCATACACTTCAAGAACAACTGGGATCTTC 720
Db      3713 TTCAAGAACCTATGCGCGACAAAATAGCATACACTTCAAGAACAACTGGGATCTTC 3772
Oy      721 TGTAGCAATTCAGAGAACTACAAATATTAACACAGCAACAACGCAACTACACATCA 780
Db      3773 TGTAGCAATTCAGAGAACTACAAATATTAACACAGCAACAACGCAACTACACATCA 3832
Oy      781 AGCAAGTCAAGATATGATGATGATGCTGTTGAAACAGCTTACAAACGCTCTTGAG 840
Db      3833 AGCAAGTCAAGATATGATGATGATGCTGTTGAAACAGCTTACAAACGCTCTTGAG 3892
Oy      841 TCAACGACATGTAGATCTGATGCTTGTCTTGTATCCAGCACAAAATCAAGTGCAG 900

```


Best Local Similarity 100.0%; Pred. No. 0;
Matches 2389; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy 1 TTCTTACGAGTTGGGACTGTATCAAGCTAGAGACGGTTAAGAAAATATCGTGTCTTA 60
Db 3053 TTCTTACGAGTTGGGACTGTATCAAGCTAGAGACGGTTAAGAAAATATCGTGTCTTA 3112
Qy 61 TATAGATGAAACAAAGCGACGCAAAAACGAGATTGTAAGTCTCTGATGAGTTAGCA 120
Db 3113 TATAGATGAAACAAAGCGACGCAAAAACGAGATTGTAAGTCTCTGATGAGTTAGCA 3172
Qy 121 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCAATCAAGATTAACAGCCAAAGCTATGTCAC 180
Db 3173 GCGTGAAGGAATCAATGCTGAGCAAAATCGTCAATCAAGATTAACAGCCAAAGCTATGTCAC 3232
Qy 181 TTTCATGAGCGACCACTATCATTTATTAACATGTTAGTTCTTATGACGCTATCATATG 240
Db 3233 TTTCATGAGCGACCACTATCATTTATTAACATGTTAGTTCTTATGACGCTATCATATG 3292
Qy 241 TGAAGAATTACTCATGAAAGATCCAAACTATAGCTAAAGATGAGATATTGTTAATGA 300
Db 3293 TGAAGAATTACTCATGAAAGATCCAAACTATAGCTAAAGATGAGATATTGTTAATGA 3352
Qy 301 GGTCAAGGGTGAATATGTTATCAAGTAGATGAAATACTATGTTTACCTTAAAGATGC 360
Db 3353 GGTCAAGGGTGAATATGTTATCAAGTAGATGAAATACTATGTTTACCTTAAAGATGC 3412
Qy 361 TGCCCAAGCGGATACGTCCTGATCAAAAAGAGAAATCAATGCACAAAACAAAGCATAG 420
Db 3413 TGCCCAAGCGGATACGTCCTGATCAAAAAGAGAAATCAATGCACAAAACAAAGCATAG 3472
Qy 421 TCACATCGTGAAGGTGGAACCTCCAAAGAAAGATGGTGTGCTTGCGACGCTTGCGCA 480
Db 3473 TCACATCGTGAAGGTGGAACCTCCAAAGAAAGATGGTGTGCTTGCGACGCTTGCGCA 3532
Qy 481 AGAAGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 540
Db 3533 AGAAGCTATACTACAGATGATGGTTATATCTTTAATGCTTCTGATATCATAGAGATAC 3592
Qy 541 TGGTGAATGCTTATATCGTCTTCTCATGAGATCATTAACATTAATCTTCTAAGATGATG 600
Db 3593 TGGTGAATGCTTATATCGTCTTCTCATGAGATCATTAACATTAATCTTCTAAGATGATG 3652
Qy 601 ATCAGCTAGCGAGTTGGCTGCTGACAGAGCCCTTCTATCTGATGAGGAAATCTGTCAA 660
Db 3653 ATCAGCTAGCGAGTTGGCTGCTGACAGAGCCCTTCTATCTGATGAGGAAATCTGTCAA 3712
Qy 661 TTCAAGAACTATATGCTGCGCAAAAATAGCATTAACATTCAGAAACAACTGGGTAACCTTC 720
Db 3713 TTCAAGAACTATATGCTGCGCAAAAATAGCATTAACATTCAGAAACAACTGGGTAACCTTC 3772
Qy 721 TGTAAAGCAATCCAGAACTAACAAATATCTAACACAAAGCAACACTTAACAGTCA 780
Db 3773 TGTAAAGCAATCCAGAACTAACAAATATCTAACACAAAGCAACACTTAACAGTCA 3832
Qy 781 AGCAAGTCAAAAGTATGATGATTAATGCTCTGAAACAGCTCTAACAACCTGCTTTAG 840
Db 3833 AGCAAGTCAAAAGTATGATGATTAATGCTCTGAAACAGCTCTAACAACCTGCTTTAG 3892
Qy 841 TCAACGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 900
Db 3893 TCAACGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 3952
Qy 901 AGCTAGAGGTGTGAGAGTCCACACGAGATCATTAACATTCATCTTCAAT 960
Db 3953 AGCTAGAGGTGTGAGAGTCCACACGAGATCATTAACATTCATCTTCAAT 4012
Qy 961 GTCTGAATTTGAAGAAGAAATGCGTATATATTCCTTGTTATGTTATGTTAAACCATG 1020
Db 4013 GTCTGAATTTGAAGAAGAAATGCGTATATATTCCTTGTTATGTTATGTTAAACCATG 4072
Qy 1021 GGTACAGATTCAGAGCCAGAAACAAAGTCCAAACCGACTCGGAACCTAGTCCAG 1080
Db 1021 GGTACAGATTCAGAGCCAGAAACAAAGTCCAAACCGACTCGGAACCTAGTCCAG 1080
```

```
Db 4073 GGTACAGATTCAGAGCCAGAAACAAAGTCCAAACCGACTCGGAACCTAGTCCAG 4132
Qy 1081 CCGGCAACCTGCACCAATCTTAAATAGACTCAAAATCTTCTTGGTATGACGCTGGT 1140
Db 4133 CCGGCAACCTGCACCAATCTTAAATAGACTCAAAATCTTCTTGGTATGACGCTGGT 4192
Qy 1141 ACGAAAAGTTGGGAGAGATATGATTCGAAGAAAAGGAGATCTCTGTTATGCTTTGC 1200
Db 4193 ACGAAAAGTTGGGAGAGATATGATTCGAAGAAAAGGAGATCTCTGTTATGCTTTGC 4252
Qy 1201 GAAAAGTTTACCATCTGAAAATCTGTTAAAATCTTGAAGCAAGTATCAAAAACAGAG 1260
Db 4253 GAAAAGTTTACCATCTGAAAATCTGTTAAAATCTTGAAGCAAGTATCAAAAACAGAG 4312
Qy 1261 TGTTCACACCTTAACTGCTAAAAAAGAAATGTTGCTCTGCTGACCAAGAAATTTTA 1320
Db 4313 TGTTCACACCTTAACTGCTAAAAAAGAAATGTTGCTCTGCTGACCAAGAAATTTTA 4372
Qy 1321 TGATTAAGCATATATATCTGTTAACTGAGGCTATTAAGCCCTGTTGNAATAAGGCTGC 1380
Db 4373 TGATTAAGCATATATATCTGTTAACTGAGGCTATTAAGCCCTGTTGNAATAAGGCTGC 4432
Qy 1381 TAATTCGATTTCCAGCCTTTAGCAAAATTTTGAACCGTTGATGATGATGATGATGAT 1440
Db 4433 TAATTCGATTTCCAGCCTTTAGCAAAATTTTGAACCGTTGATGATGATGATGATGAT 4492
Qy 1441 TAAAGAAAATTTGGTATATATTTATGCAATCTCTGACCAATTCATCCATCCAGAGCG 1500
Db 4493 TAAAGAAAATTTGGTATATATTTATGCAATCTCTGACCAATTCATCCATCCAGAGCG 4552
Qy 1501 ACTTGGCAACCAAAATCTCAAAATTTAGTATACTGAAGAGAAAGTTGCTATTT 1560
Db 4553 ACTTGGCAACCAAAATCTCAAAATTTAGTATACTGAAGAGAAAGTTGCTATTT 4612
Qy 1561 AGCTGATAGATATCAACGTCAGATGCTTACATTTTGAAGACATATATATCAAGTGA 1620
Db 4613 AGCTGATAGATATCAACGTCAGATGCTTACATTTTGAAGACATATATATCAAGTGA 4672
Qy 1621 TGAAGAGATGCAATATGTAACGCTCATATGAGGCGATAGTCACTGATGGAAGAAATAG 1680
Db 4673 TGAAGAGATGCAATATGTAACGCTCATATGAGGCGATAGTCACTGATGGAAGAAATAG 4732
Qy 1681 CTTTCTGATTAAGAAAAGTTGACGCTCAAGCCTATATTAAGAAAAGTATCTTACC 1740
Db 4733 CTTTCTGATTAAGAAAAGTTGACGCTCAAGCCTATATTAAGAAAAGTATCTTACC 4792
Qy 1741 TCCATCTCCAGAGCAGATGTTAAGCAAAATCCACTGAGATATGTCAGAGCTATTTTA 1800
Db 4793 TCCATCTCCAGAGCAGATGTTAAGCAAAATCCACTGAGATATGTCAGAGCTATTTTA 4852
Qy 1801 CAATCGTGAAGAGGGAAGAAAGCAATTCACCTGCTGCACTTCATATATGTTGAGCA 1860
Db 4853 CAATCGTGAAGAGGGAAGAAAGCAATTCACCTGCTGCACTTCATATATGTTGAGCA 4912
Qy 1861 TACAGTTGAGGTTAAAAACGGTAAATTTGATTAATCTCATTAAGATCATTAACATAAT 1920
Db 4913 TACAGTTGAGGTTAAAAACGGTAAATTTGATTAATCTCATTAAGATCATTAACATAAT 4972
Qy 1921 TAAATTTGCTTGTGATGATATCACATACAAAGCTCCAAATGCTATATCTTGGAAGA 1980
Db 4973 TAAATTTGCTTGTGATGATATCACATACAAAGCTCCAAATGCTATATCTTGGAAGA 5032
Qy 1981 TTTGTTGCGACGATTAAGTACTACGTAAGAACCCCTGAGGAAGCTCAACTTATATGA 2040
Db 5033 TTTGTTGCGACGATTAAGTACTACGTAAGAACCCCTGAGGAAGCTCAACTTATATGA 5092
Qy 2041 TGAATGGGCAATGCTCAGTACATGTTAGTGAAGCAAGAACCAAGTGAAGTCCAAA 2100
Db 5093 TGAATGGGCAATGCTCAGTACATGTTAGTGAAGCAAGAACCAAGTGAAGTCCAAA 5152
Qy 2101 TAAAGACTTCAAGCGGATGAAGGCAAGTGAAGAAACACTGCTGAGCCAGAGTCCC 2160
Db 5153 TAAAGACTTCAAGCGGATGAAGGCAAGTGAAGAAACACTGCTGAGCCAGAGTCCC 5212
```



```

/gene="SP1175"
/notes="similar to GP:4249624; identified by sequence
similarity: putative"
/codon_start=1
/transl_table=11
/product="conserved domain protein"
/protein_id="AAK75284.1"
/db_xref="GI:14972655"
/translation="MILSVCSYELGLYOARTVKEKNNRVSYIDKOATOKTENLTPDEV
SKREGINAEQIVIKITDQGYVTHSGHDHYHNGKVPYALIISEELMLPBYKIKDED
IYNEKGGYVTKDGGKYVYIKDAHANVYTKKEINQKQKSHREGGTPRDGAY
ALASQGRITTDGDTENASDI IEDTGAVTVPHGDHYVTPKELSLSELAABAFI
SGRANLSRTYRRONSNTSRITWVPSVNPNTNTNNTSNNSQASQSNIDSL
LKQIYKLPISQRHVESDGLVDPDAQITSTARGAVAPHGDIHYHPIQSOMLEBERIA
RIIPLRYSNHVPDSREQSPQPTPEPSQPOPAFMLKIDNSLVSLOEPRYKGEV
YVPEKGISRVYPAKDLPSBETIKONLESLSKQESVSHLTAKKENVAPRODFKAY
NLTTEAHKALFENKGRNSDFOALDKLRLNDESNKKEIVDDLALAPITTHEBRLG
KPSQIETDEDEVRJAQLADKRTTSBGITFEBHDITSDGAVYTPHGHSHWIGKDS
LSDEKVAQAQYTRKEGILPPSPDADVANKPTGSAALINRVKGEKIPVRLPYWY
EHTVEVKNGNLIIPKIDHYNIKEAMPDHTYKAPNGYTLDELPAITIKYVHEDEP
HSNQMGNASHEVLGKDHSDPNKPKFADEEPEVETPAEPEPOVETKEADLKEA
EVLAKVTDSLSKANATETLAGLRNNLTLQIMDNNSIWAEBKLLALLKSNPSVSK
EKIN"
/complement(8049..9782)
/gene="SP1176"
/complement(8049..9782)
/gene="SP1176"
/codon_start=1
/transl_table=11
/product="phosphoenolpyruvate-protein phosphotransferase"
/protein_id="AAK75285.1"
/db_xref="GI:14972656"
/translation="MTEMLKGIASDGVAVAKAYLLVOPDSFETITVEDINAEARL
DALOASODELSVIREKAVGTGERAAVDPALHMLVADPEMISOIKETTRAKVNE
AGLEVMTMETITPEGMEDNPMOERADIDVTRKYANILGKRLPPASINEVIV
IAHDITSDIQLDKNFYKAVVINGETSHSAIWARLILAAVLGINNIEIVKDG
ILAVNGITGEVITINTEQAAEFKAAQKAEKMAEMLLADQVTLVADGHEFLAN
IGTPRDEGVANNNGAEVAGLYRTELJYMSODFPTEDEQYAYVAIEGNMGKVVVR
TMDIGDKELPYFDMPEHMPFLGRALRISSETGDMPFQRLRALLRABVHQLRI
MPWVALIKERRAKAVDEERKANILAEGVADVNIQWIMTELPAALMADQAKYV
DPSFICTDILQYTMADRNMEQVSYIQTPNPSILRLINNVITAAAHBEGMACGE
MAGDOQAVPLVGMGLDEFMSATSVLSTRSLMKLDITAKMEVYANRALTECSTMEV
LELQKEYVNP"
/complement(9788..10051)
/gene="SP1177"
/complement(9788..10051)
/gene="SP1177"
/codon_start=1
/transl_table=11
/product="phosphocarrier protein HPr"
/protein_id="AAK75286.1"
/db_xref="GI:14972657"
/translation="MAKCDRVHVAETGTHARPATLLVOTASKRPSDITLEYKGSVNL
KSIINGVMSLIGGGADVTISAGADADDAIAISETWEKEGLA"
BASE COUNT 2842 a 2089 c 1850 g 3475 t
ORIGIN
Query Match 97.9%; Score 2338; DB 1; Length 10256;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2388; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

181 TTCACATGGCGACCATCATCATTTATCAATGTGAAGTTCCTTATGACGCTATCATCAG 240
7557 TTCCATAGGCGACCATCATCATTTATCAATGTGAAGTTCCTTATGACGCTATCATCAG 7498
241 TGAAGATTACTCATGAAAGATCCAAACTATTAAGCTAAAGATAGAGATATTTGTAATGA 300
7497 TGAAGATTACTCATGAAAGATCCAAACTATTAAGCTAAAGATAGAGATATTTGTAATGA 7438
301 GGTCAAGGGTGAATATGTTTATCAAGTAGATGAAATATCATGTTTACCTTAAGATATGC 360
7437 GGTCAAGGGTGAATATGTTTATCAAGTAGATGAAATATCATGTTTACCTTAAGATATGC 7378
361 TGCCACGCGGATACCTCCGTACAAAGAGAAATCAATGCACAAAACAAAGACATAG 420
7377 TGCCACGCGGATACCTCCGTACAAAGAGAAATCAATGCACAAAACAAAGACATAG 7318
421 TCACATCTGTGAAGTGTGAATCCAAAGAAACGATGTCCTGCTTGGACAGCTTGCCA 480
7317 TCACATCTGTGAAGTGTGAATCCAAAGAAACGATGTCCTGCTTGGACAGCTTGCCA 7258
481 AGAGCGCTATCTACAGATGATGTTTATCTTATGCTTGTGATATCATAGAGATAC 540
7257 AGAGCGCTATCTACAGATGATGTTTATCTTATGCTTGTGATATCATAGAGATAC 7198
541 TGGTATGCTTATATCGTTCCTCATGAGATCATTTACCATTAATCTTAAGATGAGTT 600
7297 TGGTATGCTTATATCGTTCCTCATGAGATCATTTACCATTAATCTTAAGATGAGTT 7138
601 ATCAGCTAGGAGTGGCTGCTGTCAGAAAGCTTCTCTATCTGCTGAGAAATCTGTCAA 660
7237 ATCAGCTAGGAGTGGCTGCTGTCAGAAAGCTTCTCTATCTGCTGAGAAATCTGTCAA 7078
661 TTCAGAACCTATGCGCCGACAAATATAGCATTAACCTTCAAGAACAACTGGGATCCTC 720
7077 TTCAGAACCTATGCGCCGACAAATATAGCATTAACCTTCAAGAACAACTGGGATCCTC 7018
721 TGTAGCAATCCAGAACTACAAATACAAACAAAGCAACAAAGCACTAACACTCA 780
7017 TGTAGCAATCCAGAACTACAAATACAAACAAAGCAACAAAGCACTAACACTCA 6958
781 AGCAAGTCAAAAGTAAATGACATTTGATGCTCTTGAACAGCTCTACAAACCTGCTTGG 840
6957 AGCAAGTCAAAAGTAAATGACATTTGATGCTCTTGAACAGCTCTACAAACCTGCTTGG 6898
841 TCAACGACATGTAGAAATCTGATGCGCTTGTCTTGAACGACAAATCAAGTCGAAC 900
6897 TCAACGACATGTAGAAATCTGATGCGCTTGTCTTGAACGACAAATCAAGTCGAAC 6838
901 AGCTAGAGTGTGACAGTGCACACGAGATCATTTACCACTTCAATCCCTTACTCAAT 960
6837 AGCTAGAGTGTGACAGTGCACACGAGATCATTTACCACTTCAATCCCTTACTCAAT 6778
961 GTCTGAATTGGAAGAAAGAAATCGCTGATATATTCCTTGTATTCCTTGTATGCTTGA 1020
6777 GTCTGAATTGGAAGAAAGAAATCGCTGATATATTCCTTGTATTCCTTGTATGCTTGA 6718
1021 GGTACAGATTCAAGGCGCAGAACCAACCAAGTCCACAAAGCTCCGAACTAGTCCAGG 1080
6717 GGTACAGATTCAAGGCGCAGAACCAACCAAGTCCACAAAGCTCCGAACTAGTCCAGG 6658
6557 CCGGCACTGACCAAAATCTTAAATATGACTCAAAATCTTCTTGTGTTAGTCACTGCT 6598
1141 ACGAAAAGTTGGGAGAGATATGATTCGAAGAAAGGCACTCTCTGTTATGCTTGGC 1200
6597 ACGAAAAGTTGGGAGAGATATGATTCGAAGAAAGGCACTCTCTGTTATGCTTGGC 6538
1201 GAAAGATTTACATCTGAAACGTTTAAATCTTGAAGCAAGTATCAAAACAAAGAGAG 1260
6537 GAAAGATTTACATCTGAAACGTTTAAATCTTGAAGCAAGTATCAAAACAAAGAGAG 6478
1261 TGTTTCACACATTTAACTGCTAAAGAAAGAAATGTTGCTCTGTCGACCAAGATTTTA 1320

```

Db	6477	TGTTTCACACCTTTAACTGCTAAAGAAAAAGAAAGTGTGCTCTGCTGACCAAGATTTTA	6418
QY	1321	TGATTAAGCATATATACTGTTAACCTGAGGCTCATTAAGCCTTGTGTTGAAATTAAGGCTG	1380
Db	6417	TGATTAAGCATATATACTGTTAACCTGAGGCTCATTAAGCCTTGTGTTGAAATTAAGGCTG	6358
QY	1381	TAATTCGATTTCCAAAGCCTTAGCAAAATTTTGTGAACGCTTGAATGATGMAATGACCTAA	1440
Db	6357	TAATTCGATTTCCAAAGCCTTAGCAAAATTTTGTGAACGCTTGAATGATGMAATGACCTAA	6298
QY	1441	TAAAGAAAAATTTGGTAGATGATTTAATGCGACTTCCTAGACCAATATCCATCCAGAGCG	1500
Db	6297	TAAAGAAAAATTTGGTAGATGATTTAATGCGACTTCCTAGACCAATATCCATCCAGAGCG	6238
QY	1501	ACTTGGCAAAACCAATTTCTCAAATTTAGTATATCGAAGACGAAGTTCGATATTCCTAAT	1560
Db	6237	ACTTGGCAAAACCAATTTCTCAAATTTAGTATATCGAAGACGAAGTTCGATATTCCTAAT	6178
QY	1561	AGCTGATPAGATPACAAAGTCAGATGCTTAATTTTGTATGTAACATATATATACATGTA	1620
Db	6177	AGCTGATPAGATPACAAAGTCAGATGCTTAATTTTGTATGTAACATATATATACATGTA	6118
QY	1621	TGAAGAGATCATATGTPAACGCTCATATGAGGCGCATATGTCAGTGGATTTGAAAAGATAG	1680
Db	6117	TGAAGAGATCATATGTPAACGCTCATATGAGGCGCATATGTCAGTGGATTTGAAAAGATAG	6058
QY	1681	CCTTTCGTATTAAGAAAAAGTTGACGCTCAAGCCTTATCTAAGAAAAAGGTATCCTACC	1740
Db	6057	CCTTTCGTATTAAGAAAAAGTTGACGCTCAAGCCTTATCTAAGAAAAAGGTATCCTACC	5998
QY	1741	TCCATCTTCCACACGCGATGTTTAAAGCAATCCATCTGAGATGTGTCAGCAGCTATTTA	1800
Db	5997	TCCATCTTCCACACGCGATGTTTAAAGCAATCCATCTGAGATGTGTCAGCAGCTATTTA	5938
QY	1801	CAATCGTGTAAGAGGGGAAAAACGAATTCACCTGGTCGATTCATATATGTTGTAACA	1860
Db	5937	CAATCGTGTAAGAGGGGAAAAACGAATTCACCTGGTCGATTCATATATGTTGTAACA	5878
QY	1861	TACAGTTAGGTTAAAAACGGTAAATTTGATTATTCCTCATPAAAGTCACTTAACATAATAT	1920
Db	5877	TACAGTTAGGTTAAAAACGGTAAATTTGATTATTCCTCATPAAAGTCACTTAACATAATAT	5818
QY	1921	TAAATTTGCTTGGTTTGATGATCAACATACAAAGCTCCAAATGGCTATACCTTGGAAAG	1980
Db	5817	TAAATTTGCTTGGTTTGATGATCAACATACAAAGCTCCAAATGGCTATACCTTGGAAAG	5758
QY	1981	TTTGTGTTGGCGCATTAAGTACTAGCTGAAACACCTCTGACGAAGTCCACATTTCTAATGA	2040
Db	5757	TTTGTGTTGGCGCATTAAGTACTAGCTGAAACACCTCTGACGAAGTCCACATTTCTAATGA	5698
QY	2041	TGATATGGGGCAATGCGCAGTGAACATGTGTTTAGGCAAGAAAGACCACTGTAAGATCCAAA	2100
Db	5697	TGATATGGGGCAATGCGCAGTGAACATGTGTTTAGGCAAGAAAGACCACTGTAAGATCCAAA	5638
QY	2101	TAAAGAACTTCAAAGCGATGGAAGACCGCATGAGGAAACACCTGCTGACCGAAGTCC	2160
Db	5637	TAAAGAACTTCAAAGCGATGGAAGACCGCATGAGGAAACACCTGCTGACCGAAGTCC	5578
QY	2161	TCAAGTAGAGCTGAAAAAGTGAAGGCCCACTCAAAGAACGAAAGTTTGTCTTGGCAA	2220
Db	5577	TCAAGTAGAGCTGAAAAAGTGAAGGCCCACTCAAAGAACGAAAGTTTGTCTTGGCAA	5518
QY	2221	AGTAAGGATTTCTAGCTGAAAGCAATGCAACGAATCTAGTGGTATTACGAAATTA	2280
Db	5517	AGTAAGGATTTCTAGCTGAAAGCAATGCAACGAATCTAGTGGTATTACGAAATTA	5458
QY	2281	TTTGAATCTTCAAATTTATGATTAACAATAGTATCATGGCAGAAAGCAGAAAAATTACTTGC	2340
Db	5457	TTTGAATCTTCAAATTTATGATTAACAATAGTATCATGGCAGAAAGCAGAAAAATTACTTGC	5398
QY	2341	GTTGTTAAAAAGAAATATCTTCACTCTGTATAGAAAAAATTAAC 2389	

RESULT 8	AX571763	349980 bp	DNA	linear	PAT 29-NOV-2002
LOCUS	AX571763/c				
DEFINITION	Sequence 4982 from Patent WO02077021.				
ACCESSION	AX571763				
VERSION	AX571763.1	GI:26003955			
KEYWORDS					
SOURCE					
ORGANISM	Streptococcus pneumoniae				
	Streptococcus pneumoniae				
	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;				
	Streptococcus.				
REFERENCE	1				
AUTHORS	Maignani,V., Tettelin,H. and Fraser,C.				
TITLE	Streptococcus pneumoniae proteins and nucleic acids				
JOURNAL	Patent: WO 02077021-A 4982 03-OCT-2002;				
	Chiron Spa (IT) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)				
FEATURES					
source	Location/Qualifiers				
	1..349980				
	/organism="Streptococcus pneumoniae"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:3313"				
	/note="seq 4979 too long; 2,162,598 bases-replaced by				
	following seq:-seq 4979: from 0,000,001 to 0,349,980-seq				
	4980: from 0,300,001 to 0,649,980-seq 4981: from 0,600,001				
	to 0,949,980-seq 4982: from 0,900,001 to 1,249,980-seq				
	4983: from 1,200,001 to 1,549,980-seq 4984: from 1,500,001				
	to 1,849,980-seq 4985: from 1,800,001 to 2,149,980-seq				
	4986: from 2,100,001 to 2,162,598"				
BASE COUNT	106896 a 71720 c 64704 g 106660 t				
ORIGIN					
Query Match	97.9%; Score 2338; DB 6; Length 349980;				
Best Local Similarity	100.0%; Pred. No. 0;				
Matches 2388; Conservative	0; Mismatches 1; Indels 0; Gaps 0;				
Qy	1	TTCTTAGCAGTGGGACCTGATCAAGCTTGAAGAAATAATCGTGTTCCTA	60		
Db	106950	TTCTTAGCAGTGGGACCTGATCAAGCTTGAAGAAATAATCGTGTTCCTA	106891		
Qy	61	TATGATGGAACAAGCAGCAAAAAACGAGATTGACTCTGATGAGTTACAA	120		
Db	106890	TATGATGGAACAAGCAGCAAAAAACGAGATTGACTCTGATGAGTTACAA	106831		
Qy	121	GCGGGAAGGATTCATGCTGAGCAAAATGCTATCAAGTAAGACCAAGGCTATGTCAC	180		
Db	106830	GCGGGAAGGATTCATGCTGAGCAAAATGCTATCAAGTAAGACCAAGGCTATGTCAC	106771		
Qy	181	TTCAATGCGCACCACCTATCATTTATTAATGATGTAAGGTTCTTATGACGATCATCAG	240		
Db	106770	TTCAATGCGCACCACCTATCATTTATTAATGATGTAAGGTTCTTATGACGATCATCAG	106711		
Qy	241	TGAAGATTACTCATGAAAGATCCAACTATPAGCTAAAGATGAGATATTGTTAATGA	300		
Db	106710	TGAAGATTACTCATGAAAGATCCAACTATPAGCTAAAGATGAGATATTGTTAATGA	106651		
Qy	301	GGTCAAGGTGATGATGTTATCAAGGTGATGGAATAATCTATGTTTACCTTAAGATGC	360		
Db	106650	GGTCAAGGTGATGATGTTATCAAGGTGATGGAATAATCTATGTTTACCTTAAGATGC	106591		
Qy	361	TGCCACGCGGATTAACGTCCTGATCAAAAGAGAAATCATGCACAAAAACAAGACATAG	420		
Db	106590	TGCCACGCGGATTAACGTCCTGATCAAAAGAGAAATCATGCACAAAAACAAGACATAG	106531		
Qy	421	TCAACATGTAAGGTGAACTCCCAAGAACGATGTCGTGTGCTTGGCAAGTTGCGCA	480		
Db	106530	TCAACATGTAAGGTGAACTCCCAAGAACGATGTCGTGTGCTTGGCAAGTTGCGCA	106471		
Qy	481	AGGACGCTATCTACAGATGATGTTATATCTTTAATGCTTGTATATCATAGAGATAC	540		
Db	106470	AGGACGCTATCTACAGATGATGTTATATCTTTAATGCTTGTATATCATAGAGATAC	106411		

QY 541 TGGTATGCTTAATATCGTTCCTCATGAGATCATTACCATTCATTCCCTAAGATGAGTT 600
 Db 106410 TGGTATGCTTAATATCGTTCCTCATGAGATCATTACCATTCATTCCCTAAGATGAGTT 106551
 QY 601 ATCAAGTACGAGTGGCTGCTGAGAAAGCCTTCATCTGCTGAGGAAATCTGTCAA 660
 Db 106350 ATCAAGTACGAGTGGCTGCTGAGAAAGCCTTCATCTGCTGAGGAAATCTGTCAA 106291
 QY 661 TTCAAGAACTATGCGCGACAAAATAAGGATTAACCTTCAGAAACAACTGGTACCTTC 720
 Db 106290 TTCAAGAACTATGCGCGACAAAATAAGGATTAACCTTCAGAAACAACTGGTACCTTC 106231
 QY 721 TGTAAAGCATCCAGAACTACAATACTACTACAAGCAACACAGCAACATAAGTCA 780
 Db 106230 TGTAAAGCATCCAGAACTACAATACTACTACAAGCAACACAGCAACATAAGTCA 106171
 QY 781 AGCAAGTCAAGATAATGACATTTAGTCTCTTGAAGAGCTTACAACTGCTTTGAG 840
 Db 106170 AGCAAGTCAAGATAATGACATTTAGTCTCTTGAAGAGCTTACAACTGCTTTGAG 106111
 QY 841 TCAAGCATATGATATGATGAGCTTGTGATCCAGACAAATCACAAGTCGAC 900
 Db 106110 TCAAGCATATGATATGATGAGCTTGTGATCCAGACAAATCACAAGTCGAC 106051
 QY 901 AGTAGAGGTGTGAGTGCACACAGAGATCATTAACCATTCATCCCTTACTCTCAAT 960
 Db 106050 AGTAGAGGTGTGAGTGCACACAGAGATCATTAACCATTCATCCCTTACTCTCAAT 105991
 QY 961 GTCTGAATTTGAGAAAGCAATGCTGTAATATCCCTTGTGATGCTTCAACCAT 1020
 Db 105990 GTCTGAATTTGAGAAAGCAATGCTGTAATATCCCTTGTGATGCTTCAACCAT 105931
 QY 1021 GGTACCATATTCAGAGCCAGAAACCAAGTCCAAACCCGAGCTCGGAACTAGTCCAG 1080
 Db 105930 GGTACCATATTCAGAGCCAGAAACCAAGTCCAAACCCGAGCTCGGAACTAGTCCAG 105871
 QY 1081 CCCGCACTGACCAAAATCTTAAATAGACTCAATCTTCTTGTGATGAGCTGCT 1140
 Db 105870 CCCGCACTGACCAAAATCTTAAATAGACTCAATCTTCTTGTGATGAGCTGCT 105811
 QY 1141 ACGAAAGTGGGAGGAGATATGATTTGAAAGAAAGGACATCTCTGTAATGCTTTC 1200
 Db 105810 ACGAAAGTGGGAGGAGATATGATTTGAAAGAAAGGACATCTCTGTAATGCTTTC 105751
 QY 1201 GAAAGATTACCTTGTGAAATCTGTTAAATCTTGAAGCAAGTTATCAAAACAAGAG 1260
 Db 105750 GAAAGATTACCTTGTGAAATCTGTTAAATCTTGAAGCAAGTTATCAAAACAAGAG 105691
 QY 1261 TGTTCACACACTTAACTGCTAAAGAAATGTTGCTCTGTCGACCAAGATTTTA 1320
 Db 105690 TGTTCACACACTTAACTGCTAAAGAAATGTTGCTCTGTCGACCAAGATTTTA 105631
 QY 1321 TGATTAAGCATATATCTGTTAAGTGAAGCTCATTAAGCCTTGTGTAAGAGGCTG 1380
 Db 105630 TGATTAAGCATATATCTGTTAAGTGAAGCTCATTAAGCCTTGTGTAAGAGGCTG 105571
 QY 1381 TAAATTCGATTTCCAGGCTTGAACAATTTATGAAGGCTTGAATGATGACCTTA 1440
 Db 105570 TAAATTCGATTTCCAGGCTTGAACAATTTATGAAGGCTTGAATGATGACCTTA 105511
 QY 1441 TAAAGAAAAATTTGATGATGATTTATTTGACCTTCTACACCAATTAACCAAGAG 1500
 Db 105510 TAAAGAAAAATTTGATGATGATTTATTTGACCTTCTACACCAATTAACCAAGAG 105451
 QY 1501 ACTTGCAAAACCAATTTCTCAATATGAGTATATGAAAGCAAGTTCGATTTGCTCAAT 1560
 Db 105450 ACTTGCAAAACCAATTTCTCAATATGAGTATATGAAAGCAAGTTCGATTTGCTCAAT 105391
 QY 1561 AGCTGATAGTATACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 1620
 Db 105390 AGCTGATAGTATACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 105331

QY 1621 TGAAGAGATGATATGTAACGCCCTCATATGGCCATAGTCACTGATTTGAAAAAGATG 1680
 Db 105330 TGAAGAGATGATATGTAACGCCCTCATATGGCCATAGTCACTGATTTGAAAAAGATG 105271
 QY 1681 CTTTTCGATTAAGAAAAAGTTGCACTCAAGCTTATCTTAAGAAAAAGGTATCTTACC 1740
 Db 105270 CTTTTCGATTAAGAAAAAGTTGCACTCAAGCTTATCTTAAGAAAAAGGTATCTTACC 105211
 QY 1741 TCCATCTCCAGACGACATGTTAAGCAATCCAACTGGAGATGATGACAGCTATTTA 1800
 Db 105210 TCCATCTCCAGACGACATGTTAAGCAATCCAACTGGAGATGATGACAGCTATTTA 105151
 QY 1801 CAATCTGCTGAAGGGGAAAAAGCAATTCACCTGTTGCACTTCATATATGTTGAGCA 1860
 Db 105150 CAATCTGCTGAAGGGGAAAAAGCAATTCACCTGTTGCACTTCATATATGTTGAGCA 105091
 QY 1861 TACAGTTGAGTTAAAAAGGTTAATTTGATTTTCTCATTAAGATCATTAATAT 1920
 Db 105090 TACAGTTGAGTTAAAAAGGTTAATTTGATTTTCTCATTAAGATCATTAATAT 105031
 QY 1921 TAAATTTGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1980
 Db 105030 TAAATTTGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 104971
 QY 1981 TTTGTTTGCACGATTAAGTACTAGTAAACACCTGACGAACGTCACATTTCTAATGA 2040
 Db 104970 TTTGTTTGCACGATTAAGTACTAGTAAACACCTGACGAACGTCACATTTCTAATGA 104911
 QY 2041 TGGATGGGGCAATGCCAGTGAAGCTGTTTGAAGCAAGAAACCAAGTGAATCCAA 2100
 Db 104910 TGGATGGGGCAATGCCAGTGAAGCTGTTTGAAGCAAGAAACCAAGTGAATCCAA 104851
 QY 2101 TAAAGCTTCAAAAGCCGATGAGAGCCAGTAAAGCAACCTGTCAGACAGAGTCCC 2160
 Db 104850 TAAAGCTTCAAAAGCCGATGAGAGCCAGTAAAGCAACCTGTCAGACAGAGTCCC 104791
 QY 2161 TCAAGTGAAGACTGAAAAAGTAAAGCCCACTCAAAAGCAAGATTTTGTTCGCAA 2220
 Db 104790 TCAAGTGAAGACTGAAAAAGTAAAGCCCACTCAAAAGCAAGATTTTGTTCGCAA 104731
 QY 2221 AGTAACGATTTCTAGTCTGAAGCCAAATGCAACGAAACTGACTGTTAGCAATTA 2280
 Db 104730 AGTAACGATTTCTAGTCTGAAGCCAAATGCAACGAAACTGACTGTTAGCAATTA 104671
 QY 2281 TTTGACTCTTCAATTAATGATTAACAATAGTATCATGACGAGAGCAAAAAATTAATTGC 2340
 Db 104670 TTTGACTCTTCAATTAATGATTAACAATAGTATCATGACGAGAGCAAAAAATTAATTGC 104611
 QY 2341 GTTGTTAAAAAGAAATATCTTCATCTGTAAGTAAAGAAAAATTAAC 2389
 Db 104610 GTTGTTAAAAAGAAATATCTTCATCTGTAAGTAAAGAAAAATTAAC 104562

RESULT 9
 AE008479/c 10320 bp DNA linear BCT 13-SEP-2001
 LOCUS
 DEFINITION Streptococcus pneumoniae R6 section 95 of 184 of the complete genome.
 ACCESSION AE008479 AE007317
 VERSION AE008479.1 GI:15458677
 KEYWORDS
 SOURCE Streptococcus pneumoniae R6
 ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Streptococcus.

REFERENCE
 AUTHORS
 1 (bases 1 to 10320)
 Hoskins, J.A., Alborn, W. Jr., Arnold, J., Blaszcak, L., Burett, S., Dehoff, B.S., Estren, S., Fritz, L., Fu, D.-J., Fuller, W., Geringer, C., Gilmour, R., Glass, J.S., Khoja, H., Kraft, A., Lagace, R., LeBlanc, D.J., Lee, L.N., Lefkowitz, E.J., Lu, J., Matsushima, P., McAhren, S., McHenry, M., McLeaster, K., Mundy, C., Niclas, T.I., Norris, F.H., O'Garra, M., Peery, R., Robertson, G.T., Rockey, P., Sun, P.-M., Winkler, M.E., Yang, Y., Young-Bellido, M., Zhao, G.,


```
/db_xref="GI:15458684"
/translation="MTEMLKGIASDVAVAKAYLLVQPLSFETITVEDTNAEARL
DALQASDELSEVIREKAVGTIGEEAAOVFAHMLVADLPMISQIKETIAKKVNAE
AGLKEVMTFITIEGMEDNPMOBRADIRDVTKRVANLGLKLPAPASINEVIVY
IADVITIGEVITINPTDQAEFKKAGSAVAKQKAEMLLDQOTVADGNHFEIAAN
IGTDPDVGANNNGAEAVGLRTEFLYMDSDPFEDDQYARAVLEGNGKEVVAR
ITMDIGDELEPYFDMPEHNPFLGRALRISISETGDMFRTOIRALRASVHQLRI
MEPVALLKEFRRAKAVPEDEKANLLAGVAVADNIQVGMIEIPAAHMLDOFAKEY
DEFSIGTNDLIOYTAADRMRNOVSLYQVNPSSLIRLNNVITKAHAEKGMACGE
MAGDQOAVPLVGMGLDEFSMSATSVLNTBSLMKLDIAKNEEVANRALTECSIMEYV
LELQKEYNVPD
complement(9861..10124)
/gene="p15h"
/ncbi="synonym: spr1063"
complement(9861..10124)
/gene="p15h"
/codon_start=1
/label_table=11
/label=SPR1063
/product="Histidine-containing phosphocARRIER protein of
the PHS"
/protein_id="AAK9867.1"
/db_xref="GI:15458685"
/translation="MASKDFHVAVETGTHARPATLLVOTASKFASDITLEYKGSVNL
KSIQVMSLQVQCADVTISAGADADDAIAAISPTMEKEGIA"
BASE COUNT      2851 a      2091 c      1862 g      3516 t
ORIGIN
Query Match      68.0%; Score 1624; DB 1; Length 10320;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 2374; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
OY      1      TTCTTACAGATTGGACCTGTATCAAGCTAAGAACGGTTAAGGAAATATCGTGTCTCTA      60
DB      7808      TTCTTACAGATTGGACCTGTATCAAGCTAAGAACGGTTAAGGAAATATCGTGTCTCTA      7749
OY      61      TATAGATGGAACAAACGACGCGCAAAAACGAGAAATTGACTCCTGTAGAGTTAGCA      120
DB      7748      TATAGATGGAACAAACGACGCGCAAAAACGAGAAATTGACTCCTGTAGAGTTAGCA      7689
OY      121      GCGTGAAGAAATCAATGCTGAGCAAAATGTCATCAAGATTAACAGACCAAGGCTATGTCAC      180
DB      7688      GCGTGAAGAAATCAATGCTGAGCAAAATGTCATCAAGATTAACAGACCAAGGCTATGTCAC      7629
OY      181      TTCACATGGCAACCACTATCATATTATTAAGTGAAGGTTCTTGAAGCTATCATAG      240
DB      7628      TTCACATGGCAACCACTATCATATTATTAAGTGAAGGTTCTTGAAGCTATCATAG      7569
OY      241      TGAAGAATTACTCATGAAGAATCCAAACTATTAAGCTAAGTGAAGATATTGTTAATGA      300
DB      7568      TGAAGAATTACTCATGAAGAATCCAAACTATTAAGCTAAGTGAAGATATTGTTAATGA      7509
OY      301      GGTCAAGGGTGAATATGTTATCAAGGTAGATGAAGAAATATATGTTTACCTTAAGGATGC      360
DB      7508      GGTCAAGGGTGAATATGTTATCAAGGTAGATGAAGAAATATATGTTTACCTTAAGGATGC      7449
OY      361      TGCCCAACCGGATACGTCCTGACAAAGAGAGAAATTCATGACGACAAAACAAAGACATAG      420
DB      7448      TGCCCAACCGGATACGTCCTGACAAAGAGAGAAATTCATGACGACAAAACAAAGACATAG      7389
OY      421      TCAACATCGTGAAGGTGGAATCCAAAGAAACGATGGTCTGTGTGCTTGCGACGTTCCCA      480
DB      7388      TCAACATCGTGAAGGTGGAATCCAAAGAAACGATGGTCTGTGTGCTTGCGACGTTCCCA      7329
OY      481      AGGACGCTATATCTACAGATGATGTTATATCTTTAATGCTTGTATCATAGAGATAC      540
DB      7328      AGGACGCTATATCTACAGATGATGTTATATCTTTAATGCTTGTATCATAGAGATAC      7269
OY      541      TGGTGAATGCTTATATCGTTCTCATGAGATCATTAACATTAATCTTAAGAAATGATTT      600
DB      7268      TGGTGAATGCTTATATCGTTCTCATGAGATCATTAACATTAATCTTAAGAAATGATTT      7209
```

```
OY      601      ATCAGTAGCGAGTTGGCTGCTGAGAGACCTTCTATCTGTGTCGAGAAATCTGTCAA      660
DB      7208      ATCAGTAGCGAGTTGGCTGCTGAGAGACCTTCTATCTGTGTCGAGAAATCTGTCAA      7149
OY      661      TTCAAGAACTTATGTCGCGACAAAATACCGATTAACATTTCAAGAACAAATCGGGTACCTTC      720
DB      7148      TTCAAGAACTTATGTCGCGACAAAATACCGATTAACATTTCAAGAACAAATCGGGTACCTTC      7089
OY      721      TGTAGCAATTCAGGAACCTCAAAATACCAACAGCAACAGCAACAACTAACATCA      780
DB      7088      TGTAGCAATTCAGGAACCTCAAAATACCAACAGCAACAGCAACAACTAACATCA      7029
OY      781      AGCAAGTCAAGTAATGATCAATGATAGTCTCTTGAACACGCTCTCAACATGCTTTGAG      840
DB      7028      AGCAAGTCAAGTAATGATCAATGATAGTCTCTTGAACACGCTCTCAACATGCTTTGAG      6969
OY      841      TCAACGACATGTATGATATCTGATGGCTTTGTCTTGTATCCAGCAAAATCAACATCGAAC      900
DB      6968      TCAACGACATGTATGATATCTGATGGCTTTGTCTTGTATCCAGCAAAATCAACATCGAAC      6909
OY      901      AGCTAGAGGTGTTGACAGTGCACACGAGATCATTTACCACTTCACTTCACTCAAT      960
DB      6908      AGCTAGAGGTGTTGACAGTGCACACGAGATCATTTACCACTTCACTTCACTCAAT      6849
OY      961      GTCGAAATTGGAAGACAAATCGCTCGTATATATCCCTTGTATGCTTCAACATTTG      1020
DB      6848      GTCGAAATTGGAAGACAAATCGCTCGTATATATCCCTTGTATGCTTCAACATTTG      6789
OY      1021      GGTACCAAGATTCAAGGCGACAGAACCAAGTCCCAACCGATCCGAACTTATCCAGG      1080
DB      6788      GGTACCAAGATTCAAGGCGACAGAACCAAGTCCCAACCGATCCGAACTTATCCAGG      6729
OY      1081      CCCGCAACCTGACCACAATCTTAAATATAGACTCAATTTCTTGTGTTAGTACGTGT      1140
DB      6728      CCCGCAACCTGACCACAATCTTAAATATAGACTCAATTTCTTGTGTTAGTACGTGT      6669
OY      1141      ACGAAAGTTGGGAGAGATATGATTTGAAAGAAAGGAGCTCTCTGTTATGCTTTTGC      1200
DB      6668      ACGAAAGTTGGGAGAGATATGATTTGAAAGAAAGGAGCTCTCTGTTATGCTTTTGC      6609
OY      1201      GAAAGATTTTACATCTGAAATCTGTTAAATATCTTGAAGCAAGTTATCAAAACAGAGAG      1260
DB      6608      GAAAGATTTTACATCTGAAATCTGTTAAATATCTTGAAGCAAGTTATCAAAACAGAGAG      6549
OY      1261      TGTTCACACACTTAACTGCTTAAAGAAATGTTGCTCTCTGAGCAACAAATTTTA      1320
DB      6548      TGTTCACACACTTAACTGCTTAAAGAAATGTTGCTCTCTGAGCAACAAATTTTA      6489
OY      1321      TGATTAAGCATATATATCTGTTAACTGAGGCTCATTAAGCCTTGTGTTGAAATTAAGGCTG      1380
DB      6488      TGATTAAGCATATATATCTGTTAACTGAGGCTCATTAAGCCTTGTGTTGAAATTAAGGCTG      6429
OY      1381      TAATTCGATTTTCCAAAGCTTTAGCAAAATTTATTAAGACGCTTGAATGATGAATGCACTTA      1440
DB      6428      TAATTCGATTTTCCAAAGCTTTAGCAAAATTTATTAAGACGCTTGAATGATGATGACCTTA      6369
OY      1441      TAAAGAAATTTGATAGATGATTTATTTGGCAATTCCTGACAAATTAACCATCCAGAGCG      1500
DB      6368      TAAAGAAATTTGATAGATGATTTATTTGGCAATTCCTGACAAATTAACCATCCAGAGCG      6309
OY      1501      ACTTGGCAACCAAAATTTCTCAATTTGATGATATCTGAAGACGAAGTTGCTGTAATTT      1560
DB      6308      ACTTGGCAACCAAAATTTCTCAATTTGATGATATCTGAAGACGAAGTTGCTGTAATTT      6249
OY      1561      AGCTGATTAAGTATCAACGCTCAGATGTTTATCTTTGATGAACATGATATTAACATGTA      1620
DB      6248      AGCTGATTAAGTATCAACGCTCAGATGTTTATCTTTGATGAACATGATATTAACATGTA      6189
OY      1621      TGAAGAGATGCAATATGTAACGCTCATATGGGCGCATAGTCACTGATTTGGAAGAAAGATAG      1680
DB      6188      TGAAGAGATGCAATATGTAACGCTCATATGGGCGCATAGTCACTGATTTGGAAGAAAGATAG      6129
OY      1681      CTTTCTGATTAAGAAAGTTGACGCTCAAGCCTATATTAAGAAAGGATATCTTACC      1740
```

QY	Db	LOCUS	DEFINITION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
QY	6128	SPNEU1915	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.	AL449937	HTG: HTGS_PHASE2.	Microb. Drug Resist.	Streptococcus pneumoniae	1	Dopazo, J., Mendoza, A., Herrero, J., Caldera, F., Humbert, Y., Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de Francesco, M., Polissi, A., Buell, G., Feger, G., Garcia, E., Peltsch, M., and
QY	1741	SPNEU1915	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.	AL449937	HTG: HTGS_PHASE2.	Microb. Drug Resist.	Streptococcus pneumoniae	2	(bases 1 to 20035)
QY	6068	SPNEU1915	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.	AL449937	HTG: HTGS_PHASE2.	Microb. Drug Resist.	Streptococcus pneumoniae	3	Dopazo, J., Mendoza, A., Herrero, J., Caldera, F., Polissi, A., Humbert, Y., Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de
QY	1801	SPNEU1915	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.	AL449937	HTG: HTGS_PHASE2.	Microb. Drug Resist.	Streptococcus pneumoniae	4	Francesco, M., Buell, G., Feger, G., Garcia, E., Peltsch, M., and
QY	6008	SPNEU1915	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.	AL449937	HTG: HTGS_PHASE2.	Microb. Drug Resist.	Streptococcus pneumoniae	5	Francesco, M., Buell, G., Feger, G., Garcia, E., Peltsch, M., and
QY	1861	SPNEU1915	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.	AL449937	HTG: HTGS_PHASE2.	Microb. Drug Resist.	Streptococcus pneumoniae	6	Francesco, M., Buell, G., Feger, G., Garcia, E., Peltsch, M., and
QY	5948	SPNEU1915	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.	AL449937	HTG: HTGS_PHASE2.	Microb. Drug Resist.	Streptococcus pneumoniae	7	Francesco, M., Buell, G., Feger, G., Garcia, E., Peltsch, M., and
QY	1921	SPNEU1915	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.	AL449937	HTG: HTGS_PHASE2.	Microb. Drug Resist.	Streptococcus pneumoniae	8	Francesco, M., Buell, G., Feger, G., Garcia, E., Peltsch, M., and
QY	5888	SPNEU1915	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.	AL449937	HTG: HTGS_PHASE2.	Microb. Drug Resist.	Streptococcus pneumoniae	9	Francesco, M., Buell, G., Feger, G., Garcia, E., Peltsch, M., and
QY	1981	SPNEU1915	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.	AL449937	HTG: HTGS_PHASE2.	Microb. Drug Resist.	Streptococcus pneumoniae	10	Francesco, M., Buell, G., Feger, G., Garcia, E., Peltsch, M., and
QY	5828	SPNEU1915	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.	AL449937	HTG: HTGS_PHASE2.	Microb. Drug Resist.	Streptococcus pneumoniae	11	Francesco, M., Buell, G., Feger, G., Garcia, E., Peltsch, M., and
QY	2041	SPNEU1915	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.	AL449937	HTG: HTGS_PHASE2.	Microb. Drug Resist.	Streptococcus pneumoniae	12	Francesco, M., Buell, G., Feger, G., Garcia, E., Peltsch, M., and
QY	5768	SPNEU1915	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.	AL449937	HTG: HTGS_PHASE2.	Microb. Drug Resist.	Streptococcus pneumoniae	13	Francesco, M., Buell, G., Feger, G., Garcia, E., Peltsch, M., and
QY	5708	SPNEU1915	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.	AL449937	HTG: HTGS_PHASE2.	Microb. Drug Resist.	Streptococcus pneumoniae	14	Francesco, M., Buell, G., Feger, G., Garcia, E., Peltsch, M., and
QY	2161	SPNEU1915	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.	AL449937	HTG: HTGS_PHASE2.	Microb. Drug Resist.	Streptococcus pneumoniae	15	Francesco, M., Buell, G., Feger, G., Garcia, E., Peltsch, M., and
QY	5648	SPNEU1915	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.	AL449937	HTG: HTGS_PHASE2.	Microb. Drug Resist.	Streptococcus pneumoniae	16	Francesco, M., Buell, G., Feger, G., Garcia, E., Peltsch, M., and
QY	2221	SPNEU1915	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.	AL449937	HTG: HTGS_PHASE2.	Microb. Drug Resist.	Streptococcus pneumoniae	17	Francesco, M., Buell, G., Feger, G., Garcia, E., Peltsch, M., and
QY	5588	SPNEU1915	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.	AL449937	HTG: HTGS_PHASE2.	Microb. Drug Resist.	Streptococcus pneumoniae	18	Francesco, M., Buell, G., Feger, G., Garcia, E., Peltsch, M., and
QY	2281	SPNEU1915	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.	AL449937	HTG: HTGS_PHASE2.	Microb. Drug Resist.	Streptococcus pneumoniae	19	Francesco, M., Buell, G., Feger, G., Garcia, E., Peltsch, M., and
QY	5528	SPNEU1915	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.	AL449937	HTG: HTGS_PHASE2.	Microb. Drug Resist.	Streptococcus pneumoniae	20	Francesco, M., Buell, G., Feger, G., Garcia, E., Peltsch, M., and
QY	2341	SPNEU1915	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.	AL449937	HTG: HTGS_PHASE2.	Microb. Drug Resist.	Streptococcus pneumoniae	21	Francesco, M., Buell, G., Feger, G., Garcia, E., Peltsch, M., and
QY	5468	SPNEU1915	Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.	AL449937	HTG: HTGS_PHASE2.	Microb. Drug Resist.	Streptococcus pneumoniae	22	Francesco, M., Buell, G., Feger, G., Garcia, E., Peltsch, M., and

FEATURES	Source	Location/Qualifiers
TITLE	Garcia-Bustos J.F.	
JOURNAL	Direct Submission	
COMMENT	Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A., Severo Ochoa 2, 28760 Tres Cantos, SPAIN * NOTE: This is a 'working draft' sequence. * This sequence will be replaced * by the finished sequence as soon as it is available and * the accession number will be preserved.	
BASE COUNT	6369 a 3635 c 4108 g 5909 t	14 others
ORIGIN		
Query Match	38.8%; Score 927; DB 2; Length 20035;	
Best Local Similarity	99.3%; Prid. No. 0;	
Matches 1497; Conservative	0; Mismatches 9; Indels 1; Gaps 1;	
Db	884	CAAATCAAGTGGAAACAGCTAGAGGTGGTGCAGTGCACACGAGATCATTTACCACTTC 943
Db	17953	CAATCAAGTGGAAACAGCTAGAGGTGGTGCAGTGCACACGAGATCATTTACCACTTC 18012
Qy	944	ATCCCTTACTCTCAAAATGTCGTAATTTGGAAGAACGAATCGCTGTATTTATTCCTTCGT 1003
Db	18013	ATCCCTTACTCTCAAAATGTCGTAATTTGGAAGAACGAATCGCTGTATTTATTCCTTCGT 18072
Qy	1004	TATCGTTCAAAACCATTTGGGTACCAAGTTCAAGGCCGAACCAACCAAGTCACACCGACT 1063
Db	18073	TATCGTTCAAAACCATTTGGGTACCAAGTTCAAGGCCGAACCAACCAAGTCACACCGACT 18132
Qy	1064	CCGGAACCTGATCGACAGGCCCGCAACCTGCACCAACTTTAAATATGACTCAAAATCTTCT 1123
Db	18133	CCGGAACCTGATCGACAGGCCCGCAACCTGCACCAACTTTAAATATGACTCAAAATCTTCT 18192
Qy	1124	TTGGTTAGTCAGCTGTGTACGAAAGTTGGGGAAGATATGATTTCCAGAAAAAGGCAATC 1183
Db	18193	TTGGTTAGTCAGCTGTGTACGAAAGTTGGGGAAGATATGATTTCCAGAAAAAGGCAATC 18252
Qy	1184	TCCTGTTATGTCTTTTGGCAAGATTTACATCTGAAACCTGTTAAAAATCTTGAAGCAAG 1243
Db	18253	TCCTGTTATGTCTTTTGGCAAGATTTACATCTGAAACCTGTTAAAAATCTTGAAGCAAG 18312
Qy	1244	TTATCAAAACAAAGAGGTGTTTCACACCTTTAATCGCTTAAAAAGAAATCTGCTCT 1303
Db	18313	TTATCAAAACAAAGAGGTGTTTCACACCTTTAATCGCTTAAAAAGAAATCTGCTCT 18372
Qy	1304	CGTCAACCAAGATTTTATGATAAGCATTAATCTGTATACGAGGCTCATTAAGCCTTG 1363
Db	18373	CGTCAACCAAGATTTTATGATAAGCATTAATCTGTATACGAGGCTCATTAAGCCTTG 18432
Qy	1364	TTTGNAAATVAGGSGTGTAAATCTGATTTTCCAAAGCCTTGAACAAATTAATTAAGCGCTTG 1423
Db	18433	TTTGNAAATVAGGSGTGTAAATCTGATTTTCCAAAGCCTTGAACAAATTAATTAAGCGCTTG 18492
Qy	1424	AATGATTAATTCGAC-TTATTAAGAAAAATTTGGTATGATTTTATTTGGCATTCCTAGCAC 1482
Db	18493	AATGATTAATTCGAC-TTATTAAGAAAAATTTGGTATGATTTTATTTGGCATTCCTAGCAC 18552
Qy	1483	AATTACCCATCCAGAGGCACTTGGCAACCAAAATTTCCAAATTTGAGTATTAAGTGAAGCA 1542
Db	18553	AATTACCCATCCAGAGGCACTTGGCAACCAAAATTTCCAAATTTGAGTATTAAGTGAAGCA 18612
Qy	1543	AGTTGTAATTCCTCAATTTAGCTGATTAAGTAAACACGTCAGATGTTACATTTTGTATGA 1602
Db	18613	AGTTGTAATTCCTCAATTTAGCTGATTAAGTAAACACGTCAGATGTTACATTTTGTATGA 18672
Qy	1603	ACATGATATAATGACGTGAACAGAGATATGATTTGAAGCCTCAATATGGGCAATATGTA 1662
Db	18673	ACATGATATAATGACGTGAACAGAGATATGATTTGAAGCCTCAATATGGGCAATATGTA 18732

```

OY 1663 CTGATTGGAAAAAGATAGCCTTTCTGATNAGAAAAAGTTGCAAGCTCAAGCCTATATCTAA 1722
Db 18733 CTGATTGGAAAAAGACAGCCTTTCTGATNAGAAAAAGTTGCAAGCTCAAGCCTATATCTAA 18792
OY 1723 AGAAAAAGTATCCCTACCTCCATCTCCCAAGCCAGATGTTAAAGCAATCCCACTGGAGA 1782
Db 18793 AGAAAAAGTATCCCTACCTCCATCTCCCAAGCCAGATGTTAAAGCAATCCCACTGGAGA 18852
OY 1783 TAGTCACAGCAGCTATTTACATCGTGTGAAAGGGAAAAACGAATTCACCTGCTGACT 1842
Db 18853 TAGTCACAGCAGCTATTTACATCGTGTGAAAGGGAAAAACGAATTCACCTGCTGACT 18912
OY 1843 TCCATATATGTTGAGCATACAGTTGAGTTAAAAACGGTAATTTGATTTCTCATPA 1902
Db 18913 TCCATATATGTTGAGCATACAGTTGAGTTAAAAACGGTAATTTGATTTCTCATPA 18972
OY 1903 GGATCATTAACCTAATATTTAAATTTGCTTGTTGATGATCAACATACAAAGCTCCAA 1962
Db 18973 GGATCATTAACCTAATATTTAAATTTGCTTGTTGATGATCAACATACAAAGCTCCAA 19032
OY 1963 TGGCTATACCTTGAAGAATTTGTTGGCAGCATTAAGTACTAGTAACACCTGAGCA 2022
Db 19033 TGGCTATACCTTGAAGAATTTGTTGGCAGCATTAAGTACTAGTAACACCTGAGCA 19092
OY 2023 ACGTCCACATTTCTAATGATGATGGGCAATGCCAGTNGCATGTGTTAGGCAGAAAGA 2082
Db 19093 ACGTCCACATTTCTAATGATGATGGGCAATGCCAGTNGCATGTGTTAGGCAGAAAGA 19152
OY 2083 CCACAGTGAAGATCCAAATTAAGAACTTCAAGCCGATGAAAGCCAGTNRAGAAAAACC 2142
Db 19153 CCACAGTGAAGATCCAAATTAAGAACTTCAAGCCGATGAAAGCCAGTNRAGAAAAACC 19212
OY 2143 TGCTGAGCCAGAAAGTCCCTCAAGTAGAGACTGAAAAAGTAAAGGCCAATCAAGAAAGC 2202
Db 19213 TGCTGAGCCAGAAAGTCCCTCAAGTAGAGACTGAAAAAGTAAAGGCCAATCAAGAAAGC 19272
OY 2203 AGAAGTTTGTCTGGGAAAGTAAACGATTTCTAGTCTGAAAGCCAAATGCAACAGAAACTCT 2262
Db 19273 AGAAGTTTGTCTGGGAAAGTAAACGATTTCTAGTCTGAAAGCCAAATGCAACAGAAACTCT 19332
OY 2263 AGCTGGTTTACGAATTAATTTGACTCTTCAATTTATGATTAACATATGATGGCAGA 2322
Db 19333 AGCTGGTTTACGAATTAATTTGACTCTTCAATTTATGATTAACATATGATGGCAGA 19392
OY 2323 AGCAGAAAAATTAATCTGCTGTTTAAAGGAAGTATCCTCATCTGTAAGTAAGAAAA 2382
Db 19393 AGCAGAAAAATTAATCTGCTGTTTAAAGGAAGTATCCTCATCTGTAAGTAAGAAAA 19452
OY 2383 AATAAAC 2389
Db 19453 AATAAAC 19459

```

RESULT 11
 SPNEU1929 702 bp, DNA linear HTG 11-JUL-2001
 LOCUS Streptococcus pneumoniae clone G54, *** SEQUENCING IN PROGRESS ***.
 ACCESSION AL449951.1 GI:11545176
 VERSION HTG; HTGS; PHASE2.
 KEYWORDS Streptococcus pneumoniae
 SOURCE Streptococcus pneumoniae
 ORGANISM Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.

REFERENCE
 1 Dopazo, J., Mendoza, A., Herrero, J., Calda, F., Humbert, Y.,
 Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de
 Francesco, M., Polissi, A., Buell, G., Feger, G., Garcia, E., Peltsch, M.
 and Garcia-Bustos, J. F.
 Annotated draft genomic sequence from a Streptococcus pneumoniae
 type 19F clinical isolate
 Microb. Drug Resist. 7 (2), 99-125 (2001)

MEDLINE 21335329
 PUBMED 11442348
 2 (bases 1 to 702)
 REFERENCE
 AUTHORS Dopazo, J., Mendoza, A., Herrero, J., Calda, F., Polissi, A.,
 Friedli, L., Guerrier, M., Grand-Schenk, E., Gandin, C., de
 Francesco, M., Buell, G., Feger, G., Garcia, E., Peltsch, M. and
 Garcia-Bustos, J. F.
 TITLE Direct Submission
 JOURNAL Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
 Severo Ochoa 2, 28760 Tres Cantos, SPAIN
 COMMENT * NOTE: This is a 'working draft' sequence.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 FEATURES
 source
 1..702
 /organism="Streptococcus pneumoniae"
 /mol_type="genomic DNA"
 /serotype="19F"
 /db_xref="taxon:1313"
 /clone="G54"
 BASE COUNT 233 a 144 c 147 g 176 t 2 others
 ORIGIN
 Query Match 13.6%; Score 326; DB 2; Length 702;
 Best Local Similarity 99.1%; Pred. No. 8.8e-164;
 Matches 576; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

OY 266 AACTATAGCTTAAAGATGAGATATGTTAATGAGTCAAGGGTATATGATATCAAG 325
Db 48 AACTATAGCTTAAAGATGAGATATGTTAATGAGTCAAGGGTATATGATATCAAG 107
OY 326 GTAGATGAAAAATATCTATGTTTACCTTAAGGATCTGCCACGGGATTAACGTCCGTACA 385
Db 108 GTAGATGAAAAATATCTATGTTTACCTTAAGGATCTGCCACGGGATTAACGTCCGTACA 167
OY 386 AAAGAGAAATCAATTCACCAAAAAACAAGACATATGTCACATCTGTAAGTGAATCCA 445
Db 168 AAAGAGAAATCAATTCACCAAAAAACAAGACATATGTCACATCTGTAAGTGAATCCA 227
OY 446 AGAAGCATGTGTGTTGCTTGGCAAGTTTGGCAAGGAGCTTACTACAGATGATGGT 505
Db 228 AGAAGCATGTGTGTTGCTTGGCAAGTTTGGCAAGGAGCTTACTACAGATGATGGT 287
OY 506 TATATCTTTAATGCTTCTGATATCATAGAGATCTGATGCTTATATGTTCTCAT 565
Db 288 TATATCTTTAATGCTTCTGATATCATAGAGATCTGATGCTTATATGTTCTCAT 347
OY 566 GGAGATCATTAACATTAATCTTAAGATGATTAATGATGAGTGGCTGCTGCA 625
Db 348 GGAGATCATTAACATTAATCTTAAGATGATTAATGATGAGTGGCTGCTGCA 407
OY 626 GAAGCTTCTCATATGTTGTTGAGGAAATCTGTCAAAATTCAGAACTTATCCGCGCAAAAT 685
Db 408 GAAGCTTCTCATATGTTGTTGAGGAAATCTGTCAAAATTCAGAACTTATCCGCGCAAAAT 467
OY 686 AGCATTAACCTTCAAGAAACAACTGGGATCTCTGTAAGCATCCAGAGATCAACAAAT 745
Db 468 AGCATTAACCTTCAAGAAACAACTGGGATCTCTGTAAGCATCCAGAGATCAACAAAT 527
OY 746 ACTAACACAAAGCAACACAGCAACATTAACAGTCAAGCAAGTCAAGTCAATGACATTGAT 805
Db 528 ACTAACACAAAGCAACACAGCAACATTAACAGTCAAGCAAGTCAAGTCAATGACATTGAT 587
OY 806 AGTCTCTTGAACAGCTCTACAAATCTGCTTTGATGATCAAG 846
Db 588 AGTCTCTTGAACAGCTCTACAAATCTGCTTTGATGATCAAG 628

```

RESULT 12
 AF340221 2535 bp, DNA linear BCT 22-MAY-2001
 LOCUS Streptococcus pneumoniae Phpa (phpa) gene, complete cds.
 DEFINITION

ACCESSION AF340221
VERSION AF340221.1 GI:13447093
KEYWORDS Streptococcus pneumoniae
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
REFERENCE 1 (bases 1 to 2535)
AUTHORS Zhang, Y., Masi, A.W., Barniak, V., Mountzouras, K., Hostetter, M.K. and Green, B.A.
TITLE Recombinant PnpA protein, a unique histidine motif-containing protein from Streptococcus pneumoniae, protects mice against intranasal pneumococcal challenge
JOURNAL Infect. Immun. 69 (6), 3827-3836 (2001)
MEDLINE 21246685
PUBMED 11349048
REFERENCE 2 (bases 1 to 2535)
AUTHORS Zhang, Y., Masi, A., Barniak, V., Mountzouras, K., Hostetter, M. and Green, B.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-2001) Department of Bacteriology, Wyeth Lederle Vaccines, 211 Bailey Road, West Henrietta, NY 14566, USA
FEATURES
source
location/Qualifiers
1..2535
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/db_xref="taxon:1313"
1..2535
/gene="pnpA"
1..2535
/note="histidine motif-containing protein"
/codon_start=1
/transl_table=11
/product="PnpA"
/protein_id="AAK26629.1"
/db_xref="GI:13447094"
/translation="MKINKYLVGSAALILSVCSYELGLYQARTVKNRRVSYIDKQATKENTLPDEVSKEKRGNAEIOIVITKTDGVTSGDGHYNGKVPYAIISSE LLMKDPNYOLKDEDIISLTKGGYVILKDGKYYVILKDAHNAVTKREKINQKQEHSHREBGTFRNDGAVVALARSQGRITTDGTFINASDITIEDGAYIVPHGDHYIIPKN ELISAEILAAKAFISLGRNLSRTRYRRQNSDNTSRITWVPSNPGTTNTNTSNNS TNSQASQNSNDISLKLQYKPLSQRHVESDGLIFDPAQITSRTANGAVPBGDHYHF IPYQSLSPLEEKLARIIPRYRSNHWVDSREQSPQSTPEPSPQAPNPQAPAS NPIDELVKEAVRKYGDGVFPENGVPRIYIPAKDIAETAAAGIDSKLAKOESLSHKL AKTDLPSDSREFVRYKADILARIHODLLDNGROVDEALDNLERLKDVSDEYKLVDDIIAFLAFLIRHPRLRGPNQITTYTDEIVAKLAGKTTEDGYIFDPRTISDEG DAYVPHMTHSHWIKKSLSEKERAQAAYAKKELTPSTDHQSGNTEAKGAAYI NRVKAARVPLDRMPYNLQYTVENKSGILIPHYHYNIKFEMDEGLYEAPKGYSL EDLATAVYVYEHPNRPHSDNGFGNASDHVORNGQADTQTEKPNBEKQTEKPE EETPEEKQSEKPEPKPTPEEPESEPESEPESEPESEPESEPESEPESEPESEPESE IONPIIKSNAKETTLGLKNNLIFGTQDNTTMAEAKLALILKESK"

BASE COUNT 880 a 518 c 533 g 604 t
ORIGIN

Query Match 8.4%; Score 200; DB 1; Length 2535;
Best Local Similarity 100.0%; Pred. No. 8.6e-96;
Matches 200; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 647 GGAATCTGTCAATTCAGAACTTATCGCGCAAAATAGCGATPAACCTTCAAGACA 706
DB 706 GGAATCTGTCAATTCAGAACTTATCGCGCAAAATAGCGATPAACCTTCAAGACA 765
QY 707 AACTGGTAACTTCTGTAGCAATCCAGAACTAACAATAACAACAACAACAAGC 766
DB 766 AACTGGTAACTTCTGTAGCAATCCAGAACTAACAATAACAACAACAACAAGC 825
QY 767 AACACTAAGTCAAGCAAGTCAAGTAATGACATTGATAGTCTCTGAAACAGGCTTAC 826
DB 826 AACACTAAGTCAAGCAAGTCAAGTAATGACATTGATAGTCTCTGAAACAGGCTTAC 885
QY 827 AAATGCTTTGAGTCAAGC 846
|||||

DB 886 AAATGCTTTGAGTCAAGC 905
RESULT 13
LOCUS AX343072 2523 bp DNA linear PAT 12-JAN-2002
DEFINITION Sequence 3 from Patent WO0198334.
ACCESSION AX343072
VERSION AX343072.1 GI:18152270
KEYWORDS
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1
AUTHORS Hamel, J., Queller, C., Charland, N., Martin, D. and Brodeur, B.
TITLE Streptococcus antigens
JOURNAL Patent: WO 0198334-A 3 27-DEC-2001;
SHIRE BIOCHEM INC. (CA)
FEATURES
source
location/Qualifiers
1..2523
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/note="BVH-11"
BASE COUNT 879 a 523 c 526 g 595 t
ORIGIN

Query Match 6.2%; Score 148; DB 6; Length 2523;
Best Local Similarity 100.0%; Pred. No. 1e-67;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 699 CAAGAACAACTGGTACTTCTGTAGCAATCCAGAACTAACAATAACAACAACA 758
DB 758 CAAGAACAACTGGTACTTCTGTAGCAATCCAGAACTAACAATAACAACAACA 817
QY 759 ACAAGAGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 818
DB 818 ACAAGAGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 877
QY 819 AGCTTAAACACTGCTTTGAGTCAAGC 846
DB 878 AGCTTAAACACTGCTTTGAGTCAAGC 905
|||||

RESULT 14
LOCUS AX343073 2647 bp DNA linear PAT 12-JAN-2002
DEFINITION Sequence 4 from Patent WO0198334.
ACCESSION AX343073
VERSION AX343073.1 GI:18152271
KEYWORDS
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1
AUTHORS Hamel, J., Queller, C., Charland, N., Martin, D. and Brodeur, B.
TITLE Streptococcus antigens
JOURNAL Patent: WO 0198334-A 4 27-DEC-2001;
SHIRE BIOCHEM INC. (CA)
FEATURES
source
location/Qualifiers
1..2647
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/note="BVH-11"
BASE COUNT 934 a 538 c 556 g 619 t
ORIGIN

Query Match 6.2%; Score 148; DB 6; Length 2647;
Best Local Similarity 100.0%; Pred. No. 1e-67;
Matches 148; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 699 CAAGAACAACTGGTACTTCTGTAGCAATCCAGAACTAACAATAACAACAACA 758
|||||

Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGATTATTCCTTCGTTATGTTCAACCAATTGGTACGATTCAAG 1035
 |||||||
 Db 944 GCTCGATTATTCCTTCGTTATGTTCAACCAATTGGTACGATTCAAG 996

RESULT 18
 AR219011
 LOCUS AR219011 2359 bp DNA linear PAT 25-SEP-2002
 DEFINITION Sequence 243 from patent US 6420135.
 ACCESSION AR219011
 VERSION AR219011.1 GI:23319945
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.

REFERENCE
 1 (bases 1 to 2359)
 Kunsch, C.A., Choi, G.H., Dillon, P.S., Rosen, C.A., Barash, S.C.,
 Fannon, M.R., and Dougherty, B.A.
 Streptococcus pneumoniae polynucleotides and sequences
 Patent: US 6420135-A 243 16-JUL-2002;
 Location/Qualifiers
 source 1..2359
 /organism="unknown"

BASE COUNT 786 a 451 c 511 g 611 t

Query Match 2.2%; Score 53; DB 6; Length 2359;
 Best Local Similarity 100.0%; Pred. No. 2e-16;
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGATTATTCCTTCGTTATGTTCAACCAATTGGTACGATTCAAG 1035
 |||||||
 Db 1879 GCTCGATTATTCCTTCGTTATGTTCAACCAATTGGTACGATTCAAG 1931

RESULT 19
 BD003923 2359 bp DNA linear PAT 31-JAN-2002
 LOCUS BD003923
 DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.
 ACCESSION BD003923.1 GI:18631884
 VERSION BD003923.1
 KEYWORDS JP 2001501833-A/243.
 SOURCE unidentified
 ORGANISM unidentified
 REFERENCE 1 (bases 1 to 2359)
 Kunsch, C.A., Choi, G.H., Dillon, P.S., Rosen, C.A., Barash, S.C.,
 Fannon, M., and Dougherty, B.A.
 Polynucleotide of Streptococcus pneumoniae and sequence
 Patent: JP 2001501833-A 243 13-FEB-2001;
 Location/Qualifiers
 OS Unidentified
 PN JP 2001501833-A/243
 PD 13-FEB-2001
 PR 30-OCT-1997 JP 1998520718
 PI 31-OCT-1996 US 60/029960
 PI CHARLES A KUNSCH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, PI
 STEVEN C BARASH,
 PI MICHAEL FANNON, BRIAN A DOUGHERTY
 PC C12N15/09, A01K67/027, C07K14/315, C07K16/12, C12N1/15, C12N1/19,
 PC C12N1/21,
 PC C12N5/10, C12P21/02, C12Q1/68, G06F17/30, C12N15/00, C12N5/00, PC
 G06F15/40
 CC Strandedness: Double;
 CC Topology: Linear;
 FH Key
 FT source 1..2359
 Location/Qualifiers
 1..2359
 /organism="unidentified"

FEATURES
 source

/mol_type="genomic DNA"
 /db_xref="taxon:32644"

BASE COUNT 786 a 451 c 511 g 611 t

QY 983 GCTCGATTATTCCTTCGTTATGTTCAACCAATTGGTACGATTCAAG 1035
 |||||||
 Db 1879 GCTCGATTATTCCTTCGTTATGTTCAACCAATTGGTACGATTCAAG 1931

RESULT 20
 AF318954
 LOCUS AF318954 2457 bp DNA linear BCT 11-FEB-2001
 DEFINITION Streptococcus pneumoniae pneumococcal histidine triad protein B
 precursor (pntB) gene, partial cds.
 ACCESSION AF318954
 VERSION AF318954.1 GI:12744741
 KEYWORDS
 SOURCE Streptococcus pneumoniae
 ORGANISM Streptococcus pneumoniae
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.

REFERENCE
 1 (bases 1 to 2457)
 Adamou, J.E., Heinrichs, J.H., Erwin, A.L., Walsh, W., Gayle, T.,
 Dormitzer, M., Kagan, R., Brehm, Y.A., Barren, P., Lathigra, R.,
 Langerman, S., Koenig, S., and Johnson, S.
 Identification and characterization of a novel family of
 pneumococcal proteins that are protective against sepsis
 Infect. Immun. 69 (2), 949-958 (2001)
 21101045

JOURNAL
 MEDLINE
 PUBMED
 11159990
 2 (bases 1 to 2457)
 Adamou, J.E., Heinrichs, J.H., Erwin, A.L., Walsh, W., Dormitzer, M., and
 Johnson, S.
 Direct Submision
 Submitted (03-NOV-2000) Molecular Microbiology, MedImmune, Inc., 35
 West Watkins Mill Road, Gaithersburg, MD 20878, USA
 Location/Qualifiers
 1..2457
 /organism="Streptococcus pneumoniae"
 /mol_type="genomic DNA"
 /serotype="4"
 /db_xref="taxon:1313"
 1..>2457
 /gene="pntB"
 1..>2457
 /gene="pntB"
 /note="pntB, protective pneumococcal antigen"
 /codon_start=1
 /transl_table=11
 /product="pneumococcal histidine triad protein B
 precursor"
 /protein_id="AAK06759.1"
 /db_xref="GI:12744742"
 /translation="MKINKYLAGSVAVALAVCSYELGRYQAGDKKESNRYAVIDG
 DOAGOKANLTPDEVSKREGINAEQIVIKITDOGVVTSHPGHHYNGVYPDAIIE
 ELMQDPNVLKSDIDVNEIKGVYIKFKNGYVLYDKAAHAAHDVIRTEERKROKER
 SHNHSRADNVVAARAGRGYTTDDGYIFNMSDIIEDGDVATVPHGHHYHPIPNEL
 SASLEAAEAAYWNGQGRSPSSSSYNNANPAQPRISRNHNLTVTPTTHQNGENISL
 LRELYAKLSEHVESDGLIFDPAQITSRITAGVAVPHGNHYH1PYEOMSELEKRIA
 R1PIPLRSNMHWVDSRPEBSPOTPEBSPBQAPNSPIDGKLKAVRVRGQGV
 FEENGVSRVYIPAKDISAETAGIDSKLAKOSLSHKLTGKTDLPSDSRPFYNKAYDL
 LAIRHODLIDKNGROVDFEALNDLIERLKQVSDSVKLVEDILAFAPLRHPRGKPK
 NAOITVTDRIQVAKLAGRYABDGIYFDPDITSDDEDAVYTPHMTSHMTKKXDSLS
 EAERAAQAYVYKEXKLTPTSPDHOSGTEAKGAAIYRNRYKAAKVPILDRMPVLOY
 TVEVNGSLITPHYHNINIKRPMFDEGLTEAPKTYTEDLLATYKYVVEHNPENPHS
 DNGFENASDHVQRNKGADVTNQTKEPSEKRPQTEKPEETPREKPSQSEKSPKT
 EEPESPESEPEQVETEKVEKLEADBLDKIOPDIKSNAKETTLGLKNLLFGT
 QDNNTIMAEKTLALKESK"

FEATURES
 source

sig_peptide 1..87
/gene="phcB"
misc_feature 247..264
/gene="phcB"
/note="Region: histidine triad"
misc_feature 616..633
/gene="phcB"
/note="Region: histidine triad"
misc_feature 718..792
/gene="phcB"
/note="Region: coiled-coil domain"
misc_feature 940..957
/gene="phcB"
/note="Region: histidine triad"
misc_feature 1396..1488
/gene="phcB"
/note="Region: coiled-coil domain"
misc_feature 1657..1674
/gene="phcB"
/note="Region: histidine triad"
misc_feature 1909..1926
/gene="phcB"
/note="Region: histidine triad"
repeat_region 525..714
/rpt_type=direct
repeat_region 1578..1758
/rpt_type=direct
BASE COUNT 832 a 796 c 533 g 593 t 3 others
ORIGIN

Query Match 2.2%; Score 53; DB 1; Length 2457;
Best Local Similarity 100.0%; Pred. No. 2e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 983 GCTGATATTATCCCTCGTTATGTTCAACCTTGGTACCGATTCAAG 1035
Db 1000 GCTGATATTATCCCTCGTTATGTTCAACCTTGGTACCGATTCAAG 1052

RESULT 21
AX569137 2457 bp DNA linear PAT 29-NOV-2002
LOCUS AX569137
DEFINITION Sequence 2345 from Patent WO02077021.
ACCESSION AX569137
VERSION AX569137.1 GI:26002635
KEYWORDS Streptococcus pneumoniae
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1
AUTHORS Maignani, V., Tettelin, H. and Fraser, C.
TITLE Streptococcus pneumoniae proteins and nucleic acids
JOURNML Patent: WO 02077021-A 2345 03-OCT-2002;
Chiron Spa (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source 1..2457
/location=Qualifiers
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/db_xref="taxon:1313"
BASE COUNT 836 a 497 c 531 g 593 t
ORIGIN

Query Match 2.2%; Score 53; DB 6; Length 2457;
Best Local Similarity 100.0%; Pred. No. 2e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 983 GCTGATATTATCCCTCGTTATGTTCAACCTTGGTACCGATTCAAG 1035
Db 1000 GCTGATATTATCCCTCGTTATGTTCAACCTTGGTACCGATTCAAG 1052

RESULT 22

AF318955
LOCUS AF318955
DEFINITION Streptococcus pneumoniae pneumococcal histidine triad protein D
ACCESSION AF318955
VERSION AF318955
KEYWORDS AF318955.1 GI:12744743
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1 (bases 1 to 2517)
AUTHORS Adamou, J. B., Heinrichs, J. H., Erwin, A. L., Walsh, W., Gayle, T.,
Dornitzer, M., Dagan, R., Brewah, Y. A., Barren, P., Lathigra, R.,
Langermann, S., Koenig, S. and Johnson, S.
TITLE Identification and characterization of a novel family of
JOURNML pneumococcal proteins that are protective against sepsis
MEDLINE Infect. Immun. 69 (2), 949-958 (2001)
PUBMED 11159990
2 (bases 1 to 2517)
REFERENCE Adamou, J. B., Heinrichs, J. H., Erwin, A. L., Walsh, W., Dornitzer, M. and
AUTHORS Johnson, S.
TITLE Direct Submission
JOURNML Submitted (03-NOV-2000) Molecular Microbiology, MedImmune, Inc., 35
West Watkins Mill Road, Gaithersburg, MD 20878, USA
FEATURES
source location=Qualifiers
1..2517
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/serotype="4"
/db_xref="taxon:1313"
1..2517
/gene="phcD"
1..2517
/gene="phcD"
/note="phcD; protective pneumococcal antigen"
/codon_start=1
/transl_table=1
/product="pneumococcal histidine triad protein D
precursor"
/protein_id="AAK06760.1"
/db_xref="GI:12744744"
/translation="MKINKKYLAGSVAVLALVSCSYELGRHOAGOVKESNRVSYIDG
DOAGKAKENLTPDEVSKREGINAEQIVITIDGVSYSHPDHVYVNGKPYDALISE
ELMKDPNYOLKSDIYNEIKGCVIVVDGKIVYLLKDAHADNIRKEIKRKQKH
SHNNGGSDQAVVAARQGRYITDDGRTFNASDITDGDATVTPYHONQENISS
LSASELAARAAVAKGKGLTPPTDQDSFTGAEAKGAELVYNNVKAQKPLDR
LRELVAKPLSERHVESDGLIFDPAQITSTRARQVAPGHNHLYPIYEQNSELEKRI
ARIIPLYRNNHWPDSPPOSPSPSPPOAPNPQAPNSPIDEKLVEAVR
KVGGVYFEENGVSRYIPAKDLSAETAGIDSKAKOESLHKGAKTDLPSDREF
YNNKAYDLARITHDLDNKRQVDFEALNDLBEIKNPDSKYLVLDIADLAPLPIH
PERGKRNQITITDDEIQVAKLAGLITPTDITDQDGTIDPRDITSDGDAYVTPHMTSH
IKKDSLEBARAAQAVAKKGLTPPTDQDSFTGAEAKGAELVYNNVKAQKPLDR
MPYNLQYTVENKQSLIIPYDHVHNKFEWFEGLAEAPRGYLEDLAVKXYVH
PNRPHSDNGFNASDHVRKKNKQDQKPEDEKSDPEPTHESDEKENHAGLNS
ADNLKRPSTDEETEEAEADTDEAIPQVENSYNNKADAEALAEKVTDPSPTRQNA
METITGKSSLLGDKNMTISAEDVSLALLKESQAPPIQ"

sig_peptide 1..87
/gene="phcD"
misc_feature 247..264
/gene="phcD"
/note="Region: histidine triad"
misc_feature 418..495
/gene="phcD"
/note="Region: coiled-coil domain"
misc_feature 619..636
/gene="phcD"
/note="Region: histidine triad"
misc_feature 943..960
/gene="phcD"
/note="Region: histidine triad"
misc_feature 1417..1509

```

misc_feature
    /gene="phtD"
    /note="Region: coiled-coil domain"
    1678..1695
    /gene="phtD"
    /note="Region: histidine triad"
    1930..1947
    /gene="phtD"
    /note="Region: histidine triad"
    526..717
    /rpt_type=direct
    repeat_region
        1597..1779
        /rpt_type=direct
    repeat_region
        509 c 545 g 633 c
    BASE COUNT
    830 a 509 c 545 g 633 c
    ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 2e-16; Length 2517;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGATTATTCCTTCGTTATCGTTCAACCAATTGGGTACCAAGATTCAAG 1035
Db 1003 GCTCGATTATTCCTTCGTTATCGTTCAACCAATTGGGTACCAAGATTCAAG 1055

RESULT 23
AX568778 2517 bp DNA linear PAT 29-NOV-2002
LOCUS AX568778
DEFINITION Sequence 1985 from Patent WO02077021.
ACCESSION AX568778
VERSION AX568778.1 GI:26002455
KEYWORDS
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 Masignani,V., Tettelin,H. and Fraser,C.
  Streptococcus pneumoniae proteins and nucleic acids
  Patent: WO 02077021-A 1985 03-OCT-2002;
  Chiron Spa (It) : THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
    source
        1..2517
        /organism="Streptococcus pneumoniae"
        /mol_type="genomic DNA"
        /db_xref="taxon:1313"
BASE COUNT
    830 a 509 c 545 g 633 c
    ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 2e-16; Length 2517;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGATTATTCCTTCGTTATCGTTCAACCAATTGGGTACCAAGATTCAAG 1035
Db 1003 GCTCGATTATTCCTTCGTTATCGTTCAACCAATTGGGTACCAAGATTCAAG 1055

RESULT 24
AX343074 2639 bp DNA linear PAT 12-JAN-2002
LOCUS AX343074
DEFINITION Sequence 5 from Patent WO0198334.
ACCESSION AX343074
VERSION AX343074.1 GI:18152272
KEYWORDS
SOURCE unidentified
ORGANISM unidentified
REFERENCE
1 Hamel,J., Ouellet,C., Charland,N., Martin,D. and Brodeur,B.
  Streptococcus antigens
  Patent: WO 0198334-A 5 27-DEC-2001;
  SHIRE BIOCHEM INC. (CA)
FEATURES
    Location/Qualifiers

```

```

source
    1..2639
    /organism="unidentified"
    /mol_type="genomic DNA"
    /db_xref="taxon:32644"
    /note="BVH-11-2"
BASE COUNT
    889 a 518 c 567 g 665 c
    ORIGIN

Query Match
Best Local Similarity 100.0%; Pred. No. 2e-16; Length 2639;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGATTATTCCTTCGTTATCGTTCAACCAATTGGGTACCAAGATTCAAG 1035
Db 1113 GCTCGATTATTCCTTCGTTATCGTTCAACCAATTGGGTACCAAGATTCAAG 1165

RESULT 25
AE007403 11931 bp DNA linear BCT 31-AUG-2001
LOCUS AE007403
DEFINITION Streptococcus pneumoniae TIGR4 section 86 of 194 of the complete genome.
ACCESSION AE007403 AE005672
VERSION AE007403.1 GI:14972469
KEYWORDS
SOURCE Streptococcus pneumoniae TIGR4
ORGANISM Streptococcus pneumoniae TIGR4
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1 (bases 1 to 11931)
  Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D.,
  Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J.,
  Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D.,
  Umayam,L.A., White,O., Salzberg,S.L., Lewis,M.R., Radune,D.,
  Holtzapple,E., Khouri,H., Wolf,A.M., Uterback,T.R., Hansen,C.L.,
  McDonald,L.A., Feldblyum,T.V., Angiuoli,S., Dickinson,T.,
  Hickey,E.K., Holt,I.E., Loftus,B.J., Yang,F., Smith,H.O.,
  Venter,J.C., Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and
  Fraser,C.M.
  Complete genome sequence of a virulent isolate of Streptococcus
  pneumoniae
  Science 293 (5529), 498-506 (2001)
  11463916
  MEDLINE
  21357209
  PUBMED
  11463916
REFERENCE
2 (bases 1 to 11931)
  Tettelin,H., Nelson,K.E., Paulsen,I.T., Eisen,J.A., Read,T.D.,
  Peterson,S., Heidelberg,J., DeBoy,R.T., Haft,D.H., Dodson,R.J.,
  Durkin,A.S., Gwinn,M., Kolonay,J.F., Nelson,W.C., Peterson,J.D.,
  Umayam,L.A., White,O., Lewis,M.R., Radune,D., Holtzapple,E.,
  Khouri,H., Wolf,A.M., Uterback,T.R., Hansen,C.L., McDonald,L.A.,
  Feldblyum,T.V., Angiuoli,S., Gesswan,P., Hickey,E.K., Holt,I.E.,
  Loftus,B.J., Uwal,M.L., Yang,F., Smith,H.O., Venter,J.C.,
  Dougherty,B.A., Morrison,D.A., Hollingshead,S.K. and Fraser,C.M.
  Direct Submission
  Submitted (29-JUN-2001) The Institute for Genomic Research, 9712
  Medical Center Dr. Rockville, MD 20850, USA
FEATURES
    source
        1..11931
        /organism="Streptococcus pneumoniae TIGR4"
        /mol_type="genomic DNA"
        /strain="TIGR4"
        /db_xref="taxon:170187"
        129..476
        /gene="SP0995"
        655..795
        /gene="SP0996"
        /note="This region contains a match to at least one other
        gene that is not full length, and is not the result of a
        sequencing artifact, similar to GP:5019553; identified by
        sequence similarity; putative; IS630-Spn1, transposase
        Orf2, truncation"
        956..1156
        /gene="SP0997"
    gene

```

```

CDS      956..1156
        /gene="SP0997"
        /note="identified by Glimmer2; putative"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="AAK75114.1"
        /db_xref="GI:14972470"
        /translation="MVASASASTSTQAOQVDSKELRALSGELDQRKALATVSDPR
        IDATKAVLDAKQABDSALTE"
        1303..1602
        /gene="SP0998"
        1303..1602
        /gene="SP0998"
        /note="identified by Glimmer2; putative"
        /codon_start=1
        /transl_table=11
        /product="hypothetical protein"
        /protein_id="AAK75115.1"
        /db_xref="GI:14972471"
        /translation="MTETNSVPAQIVVSLALGVIAFWLIRKKESEIQOLSTEL
        KVLGQIDAKQAKKVLAKQNLLQETLVKVENSGAETRTVLVEELKALIDKIK"
        1833..2540
        /gene="SP0999"
        1833..2540
        /gene="SP0999"
        /note="similar to GP:6714962; identified by sequence
        similarity; putative"
        /codon_start=1
        /transl_table=11
        /product="cytochrome c-type biogenesis protein CcdA"
        /protein_id="AAK75116.1"
        /db_xref="GI:14972472"
        /translation="METIVFLSVPLAGVLSFSPFCIPLLPVYAGILDDQESAKS
        SLFGKRLVMSGLRITLCIFAGISLRIPTLPGAGVGHILVAMRFRYOMGAIIILIGI
        HOMELFHLKLELVQKSFTRKSDSNRYNSAPLGTTFSPGRTPCIGPLVSSVLAAAS
        GGNMGMOGIVLYLITLGMALPFLVALASGLWMPYFSKIRHMMLELKIGGFLVLA
        GILLLLGVNVLVAGIFE"
        2551..3108
        /gene="SP1000"
        2551..3108
        /gene="SP1000"
        /codon_start=1
        /transl_table=11
        /product="thioredoxin family protein"
        /protein_id="AAK75117.1"
        /db_xref="GI:14972473"
        /translation="MKKVMFAGLSLSLVLMACEEETKTKTQAAQKQKTTVQOIN
        VKQADPFLQSMDEKVELSDPFRKRYLYKFWASWGPCKSKSPMLMELAKDRDRD
        ELITVIAPIQDEKTVQEPQWFOEGYKDIPLVLYATTFQVAYVRSIPTLELIDS
        OGKICKIOFGAISNDAAERAFKKN"
        3482..4873
        /gene="SP1001"
        3482..4873
        /gene="SP1001"
        /note="identified by match to PFAM protein family HM
        PF00324"
        /codon_start=1
        /transl_table=11
        /product="amino acid permease family protein"
        /protein_id="AAK75118.1"
        /db_xref="GI:14972474"
        /translation="WNIFRTKNSLDKTEMRHLKMLDILLIGAVGWGTGVPFITGG
        AAATLAPGLAVSIYIALCVGLSALPFAEPASRPVAGVASYVIALIGFPMLWAG
        MLTMEPRTAISGVASGNAAVFKGLISQYIGAPALNGTRPPQAGTVVDLLPLVLV
        LVTSLVLNAKALRFNSILVILKFSALAPLVGINIKFDNNSNFAVPGFQIYGAK
        SGVHAGSHLVMDPVALFRSVGISMAVDEKTPQKNIPRSIVLSLVTLVTLVLT
        GVLVASHNLVDPAVAFALRSVGIISWAAVSVLAILTLIVCISMTVLSMIVSLAR
        DGLVAAPELSELTSPKRNATILTLGLASAVAAGWPLASIAAFINICTLAVLIMLV
        GLTRPKRGKMKAGEFKTPVLVPLIPILSIITCISFMLQYNNMTIAETVALLVGSITII
        YFTGYKXISTIE"
        5016..5933
        /gene="SP1002"

```

[illegible]


```

NSTLEEVPTDVPQEKVAFKFAESYGMKLENVLFNMDDTIELYLPSEGEVIKKNMADFTG
EAPQNGENKSENGKVSSTGVENQPTENKPADSLPEAPNKEPVKPESTNDGMLNPE
GNVGSDDPMLDPALEAPAVDDPVQELKEFTASYGGLDSVLFNMDDTIELYLPSEGEVI
KKNLSDLIA"
gene
9403..9543
/feature="sp:0909"
CDS
9403..9543
/feature="sp:0909"
/codon_start=1
/transl_table=11
/label=sp:0909
/product="Hypothetical protein"
/protein_id="AAK9713.1"
/db_xref="GI:15458516"
/translation="MLRITLSPYQFAHILFQSTIVPHGHVHFIPESDLSAGELAVA
KV"
gene
9577..10125
/feature="pntc-truncation"
/feature="synonym: sp:0910"
9577..10125
/feature="pntc-truncation"
/codon_start=1
/transl_table=11
/label=sp:0910
/product="Pneumococcal histidine triad protein E
precursor, truncation"
/protein_id="AAK9714.1"
/db_xref="GI:15458517"
/translation="MPVTPNNGVSAVDGCVFNPNDIYRDTGDATYVHGHYIYIK
SLNNPSPHSNTEVEGSSSVLSNPSTLHVEEDDGHGFANRISDSGFIPIHG
DHNHYIKVQTKGYEALAKNKIPSLQSNYPGTFPEKAVLAKVDLLADSRISYDKPI
EQROIELALGQFTESLKKIKYS"
complement(10187..10402)
/gene="sp:0911"
complement(10187..10402)
CDS
complement(10187..10402)
/gene="sp:0911"
/codon_start=1
/transl_table=11
/label=sp:0911
/product="Hypothetical protein"
/protein_id="AAK9715.1"
/db_xref="GI:15458518"
/translation="MFQELICIAQKTFYFLAICRLLVAIYHVLKQESYNTLQGL
TEIRNPDKTMSVQDAIRFAQQGFNML"
10373..11002
/gene="sp:0912"
10373..11002
/gene="sp:0912"
/codon_start=1
/transl_table=11
/label=sp:0912
/product="Hypothetical protein"
/protein_id="AAK9716.1"
/db_xref="GI:15458519"
/translation="MSKANOLLEQIHBERGIRQSLAEVENIKAAITNKADADJDEVNSQ
VKDVIYTRIASLQOEKENAEDPQTLVLYQKLYDILMSLHLYSSNNGSDADFDKVA
LLDQSAKSDKLTALSTKALIVLNIOGKASAESETPANRANAGDTSANQP
NATASNETASDENKPSRTRSPASTSSTENKTESSTTGNGEKPEV"
complement(11045..12265)
/gene="pept"
/feature="synonym: sp:0913"
complement(11045..12265)
CDS
complement(11045..12265)
/gene="pept"
/EC_number="3.4.11.-"

```

Query Match 2.2%; Score 53; DB 1; Length 12372;
 Best Local Similarity 100.0%; Pred. No. 1.8e-16;
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 983 GCTGATATTATTCCTCGTATCGTTCAACCATTTGGGTACCGATTCAAG 1035
 |||||
 DB 3682 GCTGATATTATTCCTCGTATCGTTCAACCATTTGGGTACCGATTCAAG 3734

```

RESULT 27
SPNEU1907/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Dopazo,J., Mendoza,A., Herrero,J., Calata,F., Humbert,Y.,
Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
Francesco,M., Polissi,A., Buell,G., Feger,G., Garcia,E., Peltsch,M.
and Garcia-Bustos,J.F.
Annotated draft genomic sequence from a Streptococcus pneumoniae
type 19f clinical isolate
Microb. Drug Resist. 7 (2), 99-125 (2001)
21335329
MEDLINE
11442348
REFERENCE
AUTHORS
Dopazo,J., Mendoza,A., Herrero,J., Calata,F., Polissi,A.,
Humbert,Y., Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
Francesco,M., Buell,G., Feger,G., Garcia,E., Peltsch,M. and
Garcia-Bustos,J.F.
Direct Submission
Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
Severo Ochoa 2, 28760 Tres Cantos, SPAIN
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
Location/Qualifiers
1..75874
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/serotype="19f"
/db_xref="taxon:1313"
/clone="G54"
BASE COUNT
22476 a 15318 c 12185 g 25844 t 51 others
ORIGIN
Query Match 2.2%; Score 53; DB 2; Length 75874;  

  Best Local Similarity 100.0%; Pred. No. 1.5e-16;  

  Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

OY 983 GCTGATATTATTCCTCGTATCGTTCAACCATTTGGGTACCGATTCAAG 1035
 |||||
 DB 75161 GCTGATATTATTCCTCGTATCGTTCAACCATTTGGGTACCGATTCAAG 75109

RESULT 28
 SPNEU1901
 LOCUS
 DEFINITION
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
 Streptococcus.

REFERENCE
 AUTHORS
 Dopazo,J., Mendoza,A., Herrero,J., Calata,F., Humbert,Y.,
 Friedli,L., Guerrier,M., Grand-Schenk,E., Gandin,C., de
 Francesco,M., Polissi,A., Buell,G., Feger,G., Garcia,E., Peltsch,M.
 and Garcia-Bustos,J.F.
 Annotated draft genomic sequence from a Streptococcus pneumoniae
 type 19f clinical isolate
 Microb. Drug Resist. 7 (2), 99-125 (2001)
 21335329

```

PUBMED 11442348
REFERENCE 2 (bases 1 to 232807)
AUTHORS Dopazo, J., Mendoza, A., Herrero, J., Caldara, F., Polissi, A.,
Humbert, Y., Friedli, L., Guerrier, M., Grand-schenk, E., Gandin, C., de
Francesco, M., Buell, G., Feger, G., Garcia, E., Pettsch, M. and
Garcia-Bustos, J.F.
JOURNAL Direct Submission
Submitted (31-OCT-2000) Research Department, Glaxo Wellcome, S.A.,
Severo Ochoa 2, 28760 Tres Cantos, SPAIN
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
FEATURES
source
1..232807
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/serotype="19F"
/db_xref="taxon:1313"
/clone="G54"
BASE COUNT 70984 a 43538 c 47937 g 70253 t 95 others
ORIGIN
Query Match 2.2%; Score 53; DB 2; Length 232807;
Best Local Similarity 100.0%; Pred. No. 1.4e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGATTATTCCTTCGTTATCGTTCAACCATTTGGTACCGATTCAAG 1035
|||||
Db 201292 GCTCGATTATTCCTTCGTTATCGTTCAACCATTTGGTACCGATTCAAG 201344

RESULT 29
AX571762 349980 bp DNA linear PAT 29-NOV-2002
DEFINITION Sequence 4981 from Patent WO02077021.
ACCESSION AX571762
VERSION AX571762.1 GI:26033954
KEYWORDS Streptococcus pneumoniae
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE 1
AUTHORS Masignani, V., Tettelin, H. and Fraser, C.
TITLE Streptococcus pneumoniae proteins and nucleic acids
JOURNAL Patent: WO 02077021-A 4981 03-OCT-2002;
Chiron Spa (IT); THE INSTITUTE FOR GENOMIC RESEARCH (US)
FEATURES
source
1..349980
/organism="Streptococcus pneumoniae"
/mol_type="genomic DNA"
/db_xref="taxon:1313"
/clone="seq 4979 too long: 2.162.598 bases-replaced by
following seq:-seq 4979: from 0.000.001 to 0.349.980-seq
4980: from 0.300.001 to 0.649.980-seq 4981: from 0.600.001
to 0.949.980-seq 4982: from 0.900.001 to 1.249.980-seq
4983: from 1.200.001 to 1.549.980-seq 4984: from 1.500.001
to 1.849.980-seq 4985: from 1.800.001 to 2.149.980-seq
4986: from 2.100.001 to 2.162.598"
BASE COUNT 108592 a 61526 c 75542 g 104320 t
ORIGIN
Query Match 2.2%; Score 53; DB 6; Length 349980;
Best Local Similarity 100.0%; Pred. No. 1.4e-16;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 983 GCTCGATTATTCCTTCGTTATCGTTCAACCATTTGGTACCGATTCAAG 1035
|||||
Db 238951 GCTCGATTATTCCTTCGTTATCGTTCAACCATTTGGTACCGATTCAAG 239003

RESULT 30

```

```

AF340223
LOCUS AF340223 492 bp DNA linear SYN 22-MAY-2001
DEFINITION Synthetic construct Phpa-20 (phpa-20) gene, complete cds.
ACCESSION AF340223
VERSION AF340223.1 GI:13447097
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE 1 (bases 1 to 492)
AUTHORS Zhang, Y., Mast, A.W., Barniak, V., Mountzouras, K., Hostetter, M.K. and
Green, B.A.
TITLE Recombinant Phpa protein, a unique histidine motif-containing
protein from Streptococcus pneumoniae, protects mice against
intranasal pneumococcal challenge
JOURNAL Infect. Immun. 69 (6), 3827-3836 (2001)
MEDLINE 21246685
PUBMED 11349048
REFERENCE 2 (bases 1 to 492)
AUTHORS Zhang, Y., Mast, A., Barniak, V., Mountzouras, K., Hostetter, M. and
Green, B.
TITLE Direct Submission
JOURNAL Submitted (25-JAN-2001) Department of Bacteriology, Wyeth Lederle
Vaccines, 211 Bailey Road, West Henrietta, NY 14586, USA
FEATURES
source
1..492
/organism="synthetic construct"
/mol_type="genomic DNA"
/db_xref="taxon:32630"
/clone="derived from Streptococcus pneumoniae Phpa
sequence"
1..492
/gene="phpa-20"
1..492
/gene="phpa-20"
/feature="histidine motif-containing protein"
/codon_start=1
/transl_table=11
/product="phpa-20"
/protein_id="AAK2631.1"
/db_xref="GI:13447098"
/translation="MSSLLRLVAKPLSERHVESDGLFDPAQITRTANGAVAPHGD
HYHFPYSLPLEKLVLPFVIVOTGYQIQPEQPVHRLREPSKPAVNPQ
PAPSNPIDKLVKAVRKVGVDGVFEENGVALVYKPRILLQKQOQALIAWPSKXY
LIS"
BASE COUNT 152 a 119 c 98 g 123 t
ORIGIN
Query Match 2.1%; Score 50; DB 12; Length 492;
Best Local Similarity 100.0%; Pred. No. 9.5e-15;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 985 TCGATTATTCCTTCGTTATCGTTCAACCATTTGGTACCGATTCA 1034
|||||
Db 185 TCGATTATTCCTTCGTTATCGTTCAACCATTTGGTACCGATTCA 234

RESULT 31
AR219123 973 bp DNA linear PAT 25-SEP-2002
LOCUS AR219123
DEFINITION Sequence 355 from patent US 6420135.
ACCESSION AR219123
VERSION AR219123.1 GI:23320057
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 973)
AUTHORS Kunsch, C.A., Choi, G.H., Dillon, P.S., Rosen, C.A., Barash, S.C.,
Ramon, M.R. and Dougherty, B.A.
TITLE Streptococcus pneumoniae polynucleotides and sequences
JOURNAL Patent: US 6420135-A 355 16-JUL-2002;
Location/Qualifiers

```

```

source          1..973
                /organism="unknown"
BASE COUNT      327 a 178 c 211 g 257 t
ORIGIN
Query Match     1.8%; Score 44; DB 6; Length 973;
Best Local Similarity 100.0%; Pred. No. 1.6e-11;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 572 CATTACCATTCCTAAGATGATGATATCAGTACGAGTT 615
Db 722 CATTACCATTCCTAAGATGATGATATCAGTACGAGTT 765

RESULT 32
LOCUS BD004035 973 bp DNA linear PAT 31-JAN-2002
DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.
ACCESSION BD004035
VERSION BD004035.1 GI:18631996
KEYWORDS JP 2001501833-A/355.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 973)
AUTHORS Kunsch,C.A., Choi,G.H., Dillon,P.J., Rosen,C.A., Bara,S.C.,
TITLE Polynucleotide of Streptococcus pneumoniae and sequence
JOURNAL Patent: JP 2001501833-A 355 13-FEB-2001;
HUMAN GENOME SCIENCES INC
COMMENT OS Unidentified
PN JP 2001501833-A/355
PD 13-FEB-2001
PF 30-OCT-1997 JP 1998520718
PR 31-OCT-1996 US 60/029960
PI CHARLES A KUNSCH,GIL H CHOI,PATRICK J DILLON,CRAIG A ROSEN, PI
STEVEN C BARASH,
PI MICHAEL FANNON,BRIAN A DOUGHERTY
PC C12N15/09,A01K67/027,C07K14/315,C07K16/12,C12N1/15,C12N1/19,
PC C12N1/21,
PC C12N5/10,C12P21/02,C12Q1/68,G06F17/30,C12N15/00,C12N5/00, PC
G06F15/40
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..973 /organism='Unidentified'.
FEATURES
source Location/Qualifiers
1..973
/mol_type="genomic DNA"
/db_xref="taxon:32644"
BASE COUNT 327 a 178 c 211 g 257 t
ORIGIN
Query Match 1.8%; Score 44; DB 6; Length 973;
Best Local Similarity 100.0%; Pred. No. 1.6e-11;
Matches 44; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 572 CATTACCATTCCTAAGATGATGATATCAGTACGAGTT 615
Db 722 CATTACCATTCCTAAGATGATGATATCAGTACGAGTT 765

RESULT 33
LOCUS ARI20406 40 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 282 from patent US 6159469.
ACCESSION ARI20406
VERSION ARI20406.1 GI:14103982
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

```

```

REFERENCE Unclassified.
AUTHORS 1 (bases 1 to 40)
        Choi,G.H., Kunsch,C.A., Barash,S.C., Dillon,P.J., Dougherty,B.,
        Fannon,M.R. and Rosen,C.A.
TITLE Streptococcus pneumoniae antigens and vaccines
JOURNAL Patent: US 6159469-A 282 12-DEC-2000;
FEATURES Location/Qualifiers
source 1..40
BASE COUNT 10 a 6 c 7 g 17 t
ORIGIN
Query Match 1.3%; Score 30; DB 6; Length 40;
Best Local Similarity 100.0%; Pred. No. 0.00073;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2360 CCTTCATCTGTAACTAAGAAAAATAAAC 2389
Db 40 CCTTCATCTGTAACTAAGAAAAATAAAC 11

RESULT 34
LOCUS BD063415 40 bp DNA linear PAT 27-AUG-2002
DEFINITION Streptococcus pneumoniae antigens and vaccines.
ACCESSION BD063415
VERSION BD063415.1 GI:22609018
KEYWORDS JP 2001505415-A/169.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 40)
AUTHORS Kunsch,C.A., Choi,G.H., Johnson,S.L. and Hromockyj,A.
TITLE Streptococcus pneumoniae antigens and vaccines
JOURNAL Patent: JP 2001505415-A 169 24-APR-2001;
HUMAN GENOME SCIENCES INC
COMMENT OS Unidentified
PN JP 2001505415-A/169
PD 24-APR-2001
PF 30-OCT-1997 JP 1998520667
PR 31-OCT-1996 US 60/029960
PI CHARLES A KUNSCH,GIL H CHOI,SYNOR L JOHNSON,ALEX HROMOCKYJ PC
C12N15/31,C12N5/18,C12N1/21,C07K14/315,C12Q1/68,A61K39/09, PC
G01N33/569,
PC G01N33/68
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FEATURES
source Location/Qualifiers
1..40
/mol_type="genomic DNA"
/db_xref="taxon:32644"
BASE COUNT 10 a 6 c 7 g 17 t
ORIGIN
Query Match 1.3%; Score 30; DB 6; Length 40;
Best Local Similarity 100.0%; Pred. No. 0.00073;
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2360 CCTTCATCTGTAACTAAGAAAAATAAAC 2389
Db 40 CCTTCATCTGTAACTAAGAAAAATAAAC 11

RESULT 35
LOCUS ARI20405 37 bp DNA linear PAT 16-MAY-2001
DEFINITION Sequence 281 from patent US 6159469.
ACCESSION ARI20405
VERSION ARI20405.1 GI:14103981
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.

```

Unclassified.
REFERENCE 1 (bases 1 to 37)
AUTHORS Choi, G.H., Kunsch, C.A., Barash, S.C., Dillon, P.J., Dougherty, B.,
TITLE Streptococcus pneumoniae antigens and vaccines
JOURNAL Patent: US 6159469-A 2812-DEC-2000;
FEATURES Location/Qualifiers
source 1..37
BASE COUNT 8 a 8 c 10 g 11 t
ORIGIN

Query Match 1.1%; Score 27; DB 6; Length 37;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTCTTACGAGTTGGAGCTGTATCAAGC 27
Db 11 TTCTTACGAGTTGGAGCTGTATCAAGC 37

RESULT 36
BD063414 37 bp DNA linear PAT 27-AUG-2002
LOCUS BD063414
DEFINITION Streptococcus pneumoniae antigens and vaccines.
ACCESSION BD063414.1 GI:22609017
VERSION JP 2001505415-A/168.
KEYWORDS unclassified
SOURCE unclassified
ORGANISM unclassified
REFERENCE 1 (bases 1 to 37)
AUTHORS Kunsch, C.A., Choi, G.H., Johnson, S.L. and Hromocky, A.
TITLE Streptococcus pneumoniae antigens and vaccines
JOURNAL Patent: JP 2001505415-A 168 24-APR-2001;
COMMENT HUMAN GENOME SCIENCES INC
PN JP 2001505415-A/168
PD 24-APR-2001
PF 30-OCT-1997 JP 1998520667
PR 31-OCT-1996 US 60/029960
PI CHARLES A KUNSCH, GIL H CHOI, SYDNOR L JOHNSON, ALEX HROMOCKY J PC
C12N15/31, C12N5/18, C12N1/21, C07K14/315, C12Q1/68, A61K39/09, PC
G01N33/569,
PC G01N33/68
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FEATURES Location/Qualifiers
source 1..37
/organism="unclassified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
BASE COUNT 8 a 8 c 10 g 11 t
ORIGIN

Query Match 1.1%; Score 27; DB 6; Length 37;
Best Local Similarity 100.0%; Pred. No. 0.03;
Matches 27; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TTCTTACGAGTTGGAGCTGTATCAAGC 27
Db 11 TTCTTACGAGTTGGAGCTGTATCAAGC 37

RESULT 37
AX343114 33 bp DNA linear PAT 12-JAN-2002
LOCUS AX343114
DEFINITION Sequence 45 from Patent WO0198334.
ACCESSION AX343114
VERSION AX343114.1 GI:18152294
KEYWORDS unclassified
SOURCE unclassified
ORGANISM unclassified

Unclassified.
REFERENCE 1
AUTHORS Hamel, J., Ouellet, C., Charland, N., Martin, D. and Brodeur, B.
TITLE Streptococcus antigens
JOURNAL Patent: WO 0198334-A 45 27-DEC-2001;
SHIRE BIOCHEM INC. (CA)
FEATURES Location/Qualifiers
source 1..33
/organism="unclassified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/note="HAMJ 282"
BASE COUNT 7 a 9 c 7 g 10 t
ORIGIN

Query Match 1.0%; Score 25; DB 6; Length 33;
Best Local Similarity 100.0%; Pred. No. 0.37;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 998 CTTCGTATCGTTCAACCATTTGG 1022
Db 9 CTTCGTATCGTTCAACCATTTGG 33

RESULT 38
AX343116 34 bp DNA linear PAT 12-JAN-2002
LOCUS AX343116
DEFINITION Sequence 47 from Patent WO0198334.
ACCESSION AX343116
VERSION AX343116.1 GI:18152296
KEYWORDS unclassified
SOURCE unclassified
ORGANISM unclassified
REFERENCE 1
AUTHORS Hamel, J., Ouellet, C., Charland, N., Martin, D. and Brodeur, B.
TITLE Streptococcus antigens
JOURNAL Patent: WO 0198334-A 47 27-DEC-2001;
SHIRE BIOCHEM INC. (CA)
FEATURES Location/Qualifiers
source 1..34
/organism="unclassified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/note="HAMJ 284"
BASE COUNT 10 a 8 c 7 g 9 t
ORIGIN

Query Match 1.0%; Score 25; DB 6; Length 34;
Best Local Similarity 100.0%; Pred. No. 0.37;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 797 GACATGATAGTCTCTTGAACAGC 821
Db 10 GACATGATAGTCTCTTGAACAGC 34

RESULT 39
ARI20328 1342 bp DNA linear PAT 16-MAY-2001
LOCUS ARI20328
DEFINITION Sequence 181 from patent US 6159469.
ACCESSION ARI20328
VERSION ARI20328.1 GI:14103904
KEYWORDS Unknown.
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 1342)
AUTHORS Choi, G.H., Kunsch, C.A., Barash, S.C., Dillon, P.J., Dougherty, B.,
TITLE Streptococcus pneumoniae antigens and vaccines
JOURNAL Patent: US 6159469-A 181 12-DEC-2000;
FEATURES Location/Qualifiers

source 1. .1342
/organism="unknown"
BASE COUNT 451 a 270 c 261 g 360 t
ORIGIN

Query Match 1.0%; Score 24; DB 6; Length 1342;
Best Local Similarity 100.0%; Pred. No. 0.96;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 546 ATGCTTATATCGTTCCTCATGAG 569
|||||
525 ATGCTTATATCGTTCCTCATGAG 548

RESULT 40
BD063337 1342 bp DNA linear PAT 27-AUG-2002
LOCUS Streptococcus pneumoniae antigens and vaccines.
ACCESSION BD063337
VERSION BD063337.1 GI:22608940
KEYWORDS JP 2001505415-A/91.
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1 (bases 1 to 1342)
AUTHORS Kunsch,C.A., Choi,G.H., Johnson,S.L. and Hromockyj,A.
TITLE Streptococcus pneumoniae antigens and vaccines
JOURNAL Patent: JP 2001505415-A 91 24-APR-2001;
HUMAN GENOME SCIENCES INC
PN JP 2001505415-A/91
PD 24-APR-2001
PF 30-OCT-1997 JP 1998520667
PI 31-OCT-1996 US 60/029960
PI CHARLES A KUNSCH,GIL H CHOI,SYDOR L JOHNSON,ALEX HROMOCKYJ PC
CI 12N15/31,C12N5/18,C12N1/21,C07K14/315,C12Q1/68,A61K39/09, PC
G01N33/569,
PC G01N33/68
CC Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers.

FEATURES
source 1. .1342
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"

BASE COUNT 451 a 270 c 261 g 360 t
ORIGIN

Query Match 1.0%; Score 24; DB 6; Length 1342;
Best Local Similarity 100.0%; Pred. No. 0.96;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 546 ATGCTTATATCGTTCCTCATGAG 569
|||||
525 ATGCTTATATCGTTCCTCATGAG 548

RESULT 41
AX608407 1446 bp DNA linear PAT 17-FEB-2003
LOCUS Sequence 6336 from Patent WO02092818.
ACCESSION AX608407
VERSION AX608407.1 GI:28403945
KEYWORDS Streptococcus agalactiae
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.

REFERENCE 1
AUTHORS Glaser,P., Rusniok,C., Chevalier,F., Frangoul,L., Laloui,L.,
Zouine,M., Couve,E., Buchrieser,C., Poyart,C., Tilen-Cuot,P. and
Kunst,F.
TITLE Streptococcus agalactiae genome sequence, use for developing

JOURNAL vaccines, diagnostic tools, and for identifying therapeutic targets
Patent: WO 02092818-A 6336 21-NOV-2002;
INSTITUT PASTEUR (PR) ; CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (PR)

FEATURES
source 1. .1446
/organism="Streptococcus agalactiae"
/mol_type="genomic DNA"
/db_xref="taxon:1311"

BASE COUNT 520 a 279 c 280 g 367 t
ORIGIN

Query Match 1.0%; Score 24; DB 6; Length 1446;
Best Local Similarity 100.0%; Pred. No. 0.96;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 485 CGCTATCTACAGATGATGTTAT 508
|||||
583 CGCTATCTACAGATGATGTTAT 606

RESULT 42
AX343078 2528 bp DNA linear PAT 12-JAN-2002
LOCUS Sequence 9 from Patent WO0198334.
ACCESSION AX343078
VERSION AX343078.1 GI:18152273
KEYWORDS unidentified
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1
AUTHORS Hamel,J., Ouellet,C., Charland,N., Martin,D. and Brodeur,B.
TITLE Streptococcus antigens
JOURNAL Patent: WO 0198334-A 9 27-DEC-2001;
SHIRE BIOCHEM INC. (CA)
LOCATION/Qualifiers

FEATURES
source 1. .2528
/organism="unidentified"
/mol_type="genomic DNA"
/db_xref="taxon:32644"
/note="BVH-3"

BASE COUNT 960 a 398 c 505 g 665 t
ORIGIN

Query Match 1.0%; Score 24; DB 6; Length 2528;
Best Local Similarity 100.0%; Pred. No. 0.92;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 546 ATGCTTATATCGTTCCTCATGAG 569
|||||
536 ATGCTTATATCGTTCCTCATGAG 559

RESULT 43
AX568780 3117 bp DNA linear PAT 29-NOV-2002
LOCUS Sequence 1987 from Patent WO02077021.
ACCESSION AX568780
VERSION AX568780.1 GI:26002456
KEYWORDS Streptococcus pneumoniae
SOURCE Streptococcus pneumoniae
ORGANISM Streptococcus pneumoniae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.

REFERENCE 1
AUTHORS Masignani,V., Tettelin,H. and Fraser,C.
TITLE Streptococcus pneumoniae proteins and nucleic acids
JOURNAL Patent: WO 02077021-A 1987 03-OCT-2002;
Chiron Spa (IT) ; THE INSTITUTE FOR GENOMIC RESEARCH (US)
LOCATION/Qualifiers

FEATURES
source 1. .3117
/organism="Streptococcus pneumoniae"

```

BASE COUNT      1150 a      526 c      612 g      829 t
ORIGIN
Query Match      1.0%; Score 24; DB 6; Length 3117;
Best Local Similarity 100.0%; Pred. No. 0.9;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      546 ATGCTATATGCTTCCTCATGAG 569
      |||||
      596 ATGCTATATGCTTCCTCATGAG 619

RESULT 44
AF318956      3120 bp      DNA      linear      BCT 11-FEB-2001
DEFINITION    Streptococcus pneumoniae pneumococcal histidine triad protein E
ACCESSION     AF318956
VERSION       AF318956.1 GI:12744745
KEYWORDS
SOURCE        Streptococcus pneumoniae
ORGANISM      Streptococcus pneumoniae
REFERENCE     1 (bases 1 to 3120)
AUTHORS       Adamou,J.E., Heinrichs,J.H., Erwin,A.L., Walsh,W., Dormitzer,M. and
              Johnson,S.
TITLE         Direct Submission
JOURNAL       Submitted (03-NOV-2000) Molecular Microbiology, MedImmune, Inc., 35
              West Watkins Mill Road, Gaithersburg, MD 20878, USA
FEATURES
  source
    1..3120
      /organism="Streptococcus pneumoniae"
      /mol_type="genomic DNA"
      /seqtype="g"
      /db_xref="taxon:1313"
    1..3120
      /gene="phtE"
    1..3120
      /gene="phtE"
      /note="phtE, protective pneumococcal antigen"
      /codon_start=1
      /transl_table=1
      /product="pneumococcal histidine triad protein E precursor"
      /protein_id="AAK06761.1"
      /db_xref="GI:12744746"
      /translation="MKFSKRYIAAGSAVIVSLCAVALNHRQENKDNKNVSVYDG
      SOSQSKSENLTPDOVSQKEIGAEQIVITIGQGVTSIGDHYHYNGVVPDALFSE
      ELMKDPNYQLKDAIDVNEKGGYIIVKQGYVYLKDAHADNVRATKDEINRQEH
      VKDNEKSNVAVARSQGRYTNQGVNPNADIIDTGNAYIVPHGHHYVPRKSDLS
      ASGLAAKAHLAGKMQBQSLSYSTASNNQTSVAKSGTSPKASLEKSLKELY
      DSPSKQRTSESDGLVFDPAKTIISRTNGVAIFGHGHHFTPYSKLSALEKIAVNPVI
      SGTGSTSNAPKNEVSVLSGLSSNPSSLTSPSSASDGYIFNPKDIVEITAY
      IVRGHDHPHYIPKSNQIGQPTLPNNSLATPSPSLINPQTSHEKHEEDGFDARII
      ADESGFVMSHGDHNYFPKDLTEBOIKAAQKHLLEVQTSNGHLSHSDQDVPN
      AKEMKDLDEKIEKAGIMKQYGVRESIVVAKENRAIIVPHGHHADPIDENHPVG
      IGHSNVEELFKPEGVAKKGNKYTGELNVNVLKNSPNNONFTLANGQKRS
      FSPPELEKGLGINLVKILITPDGYLEKVSQKVGEGVNIANBELDPIYLPQGTFFK
      YTIASKDYPEVSVDGTFVPTSLAYKMAQTIIFYPHADYIARVNPQAVPRGDAL
      VVAFDEPHGNALNNYKVELIKLPKLNQGTITAGNKILVTNPMANVLDNOSTYI
      VEVPILEKENQTDKPSILPQFRKNAQENLIKDEKVEEPKTSSEKYEKELSETGNST

```

```

NSTLEEVPTVDVQEKVAKFAESYGMKLENVLFNMDGTIELYLPSEGEVIKKQMDPTG
EAPQNGENKPSKSENGKSVSTGTVENOPTENKPADSLPEAPNPKPVKPESTNDGMNPE
GNVSGDPEMLDPALEBAPAVDVQEKLEKFTASYGLGDSVIFNMDGTIELRNPSEGEVI
KKNSDLIA"
      1..87
      /gene="phtE"
      247..264
      /gene="phtE"
      /note="Region: histidine triad"
      442..531
      /gene="phtE"
      /note="Region: coiled-coil domain"
      613..630
      /gene="phtE"
      /note="Region: histidine triad"
      925..942
      /gene="phtE"
      /note="Region: histidine triad"
      1186..1203
      /gene="phtE"
      /note="Region: histidine triad"
      1381..1398
      /gene="phtE"
      /note="Region: histidine triad"
      1414..1491
      /gene="phtE"
      /note="Region: coiled-coil domain"
      1648..1665
      /gene="phtE"
      /note="Region: histidine triad"
      522..789
      /rpt_type=direct
      1569..1749
      /rpt_type=direct

BASE COUNT      1152 a      526 c      612 g      830 t
ORIGIN
Query Match      1.0%; Score 24; DB 1; Length 3120;
Best Local Similarity 100.0%; Pred. No. 0.9;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      546 ATGCTATATGCTTCCTCATGAG 569
      |||||
      596 ATGCTATATGCTTCCTCATGAG 619

RESULT 45
AX343070      3120 bp      DNA      linear      PAT 12-JAN-2002
LOCUS         AX343070
DEFINITION    Sequence 1 from Patent WO0198334.
ACCESSION     AX343070
VERSION       AX343070.1 GI:18152268
KEYWORDS
SOURCE        unidentified
ORGANISM      unidentified
REFERENCE     1
AUTHORS       Hamel,J., Ouellet,C., Charland,N., Martin,D. and Brodeur,B.
TITLE         Streptococcus antigens
JOURNAL       Patent: WO 0198334-A 1 27-DEC-2001;
              SHIRE BIOCHEM INC. (CA)
FEATURES
  source
    1..3120
      /organism="unidentified"
      /mol_type="genomic DNA"
      /db_xref="taxon:32644"
      /note="BVH-3"

BASE COUNT      1152 a      525 c      612 g      831 t
ORIGIN
Query Match      1.0%; Score 24; DB 6; Length 3120;
Best Local Similarity 100.0%; Pred. No. 0.9;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 546 ATGCTTATATCGTTCCTCATGAG 569
 |||||
 Db 596 ATGCTTATATCGTTCCTCATGAG 619

RESULT 46
 AX343071 5048 bp DNA linear PAT 12-JAN-2002
 LOCUS AX343071
 DEFINITION Streptococcus pneumoniae 2 from Patent WO0198334.
 ACCESSION AX343071
 VERSION AX343071.1 GI:18152269
 KEYWORDS
 SOURCE unidentified
 ORGANISM Streptococcus pneumoniae
 unclassified.

REFERENCE 1
 AUTHORS Hamel,J., Ouellet,C., Charland,N., Martin,D. and Brodeur,B.
 TITLE Streptococcus antigens
 JOURNAL Patent: WO 0198334-A 2 27-DEC-2001;
 SHIRE BIOCHEM INC. (CA)
 FEATURES
 source Location/Qualifiers
 1..5048
 /organism="unidentified"
 /mol_type="genomic DNA"
 /db_xref="taxon:32644"
 /note="BVH-3"

BASE COUNT 1709 a 907 c 1104 g 1328 t
 ORIGIN

Query Match 1.0%; Score 24; DB 6; Length 5048;
 Best Local Similarity 100.0%; Pred. No. 0.87;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 546 ATGCTTATATCGTTCCTCATGAG 569
 |||||
 Db 2372 ATGCTTATATCGTTCCTCATGAG 2395

RESULT 47
 AR218960/c 6867 bp DNA linear PAT 25-SEP-2002
 LOCUS AR218960
 DEFINITION Sequence 192 from Patent US 6420135.
 ACCESSION AR218960
 VERSION AR218960.1 GI:23319894
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Streptococcus pneumoniae
 unclassified.

REFERENCE 1 (bases 1 to 6867)
 AUTHORS Kunsch,C.A., Choi,G.H., Dillon,P.S., Rosen,C.A., Barash,S.C.,
 Fannon,M.R. and Dougherty,B.A.
 TITLE Streptococcus pneumoniae polynucleotides and sequences
 JOURNAL Patent: US 6420135-A 192 16-JUL-2002;
 FEATURES
 source Location/Qualifiers
 1..6867
 /organism="unknown"

BASE COUNT 1896 a 1325 c 1212 g 2433 t
 ORIGIN

Query Match 1.0%; Score 24; DB 6; Length 6867;
 Best Local Similarity 100.0%; Pred. No. 0.85;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 546 ATGCTTATATCGTTCCTCATGAG 569
 |||||
 Db 6190 ATGCTTATATCGTTCCTCATGAG 6167

RESULT 48
 BD003872/c 6867 bp DNA linear PAT 31-JAN-2002
 LOCUS BD003872
 DEFINITION Polynucleotide of Streptococcus pneumoniae and sequence.

ACCESSION BD003872
 VERSION BD003872.1 GI:18631833
 KEYWORDS JP 2001501833-A/192.
 SOURCE unidentified
 ORGANISM Streptococcus pneumoniae
 unclassified.

REFERENCE 1 (bases 1 to 6867)
 AUTHORS Kunsch,C.A., Choi,G.H., Dillon,P.J., Rosen,C.A., Barash,S.C.,
 Fannon,M. and Dougherty,B.A.
 TITLE Polynucleotide of Streptococcus pneumoniae and sequence
 JOURNAL Patent: JP 2001501833-A 192 13-FEB-2001;
 HUMAN GENOME SCIENCES INC
 COMMENT
 OS Unidentified
 PN JP 2001501833-A/192
 PD 13-FEB-2001
 PR 30-OCT-1997 JP 1998520718
 PI 31-OCT-1996 US 60/029960
 PI CHARLES A KUNSCH, GIL H CHOI, PATRICK J DILLON, CRAIG A ROSEN, PI
 STEVEN C BARASH,
 MICHAEL FANNON, BRIAN A DOUGHERTY
 PC C12N15/09, A01K67/027, C07K14/315, C07K16/12, C12N1/15, C12N1/19,
 PC C12N1/21,
 PC C12N5/10, C12P21/02, C12Q1/68, G06F17/30, C12N15/00, C12N5/00, PC
 G06F15/40
 CC Strandedness: Double;
 CC Topology: Linear;
 FH Key
 FT source Location/Qualifiers
 1..6867
 /organism="unidentified"
 /mol_type="genomic DNA"
 /db_xref="taxon:32644"

BASE COUNT 1896 a 1325 c 1212 g 2433 t
 ORIGIN

Query Match 1.0%; Score 24; DB 6; Length 6867;
 Best Local Similarity 100.0%; Pred. No. 0.85;
 Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 546 ATGCTTATATCGTTCCTCATGAG 569
 |||||
 Db 6190 ATGCTTATATCGTTCCTCATGAG 6167

RESULT 49
 AE014279 21397 bp DNA linear BCT 23-SEP-2002
 LOCUS AE014279/c
 DEFINITION Streptococcus agalactiae 2603V/R section 89 of 100 of the complete
 genome.
 ACCESSION AE014279 AE009948
 VERSION AE014279.1 GI:22534974
 KEYWORDS
 SOURCE Streptococcus agalactiae 2603V/R
 ORGANISM Streptococcus agalactiae 2603V/R
 unclassified.

REFERENCE 1 (bases 1 to 21397)
 AUTHORS Tettelin,H., Masiugnani,V., Cieslewicz,M.J., Eisen,J.A.,
 Tettelin,H., Masiugnani,V., Cieslewicz,M.J., Eisen,J.A.,
 Peterson,S., Wessels,M.R., Paulsen,I.T., Nelson,K.E., Margarit,I.,
 Read,T.D., Madoff,L.C., Wolf,A.M., Beaman,M.J., Brinkac,L.M.,
 Dougherty,S.C., DebRoy,R.T., Durkin,S., Kojonay,J.F., Umayam,L.A.,
 Madupu,R., Lewis,M.R., Radune,D., Fedorova,N.B., Scanlan,D.,
 Khouiri,H., Mulligan,S., Carly,H.A., Cline,R.T., Gill,J.,
 Sarselli,M., Mora,M., Iacobini,E.T., Brettoni,C., Galli,G.,
 Mariani,M., Vegni,F., Malone,D., Rinaudo,D., Rappuoli,R.,
 Telford,J.L., Kasper,D.L., Grandi,G. and Fraser,C.M.
 TITLE Complete genome sequence and comparative genomic analysis of an
 emerging human pathogen, serotype V Streptococcus agalactiae
 Proc. Natl. Acad. Sci. U.S.A. 99 (19), 12391-12396 (2002)
 PUBMED 12200547
 REFERENCE 2 (bases 1 to 21397)

AUTHORS

Tettein, H., Masignani, V., Cieslewicz, M. J., Eisen, J. A., Peterson, S., Wessels, M. R., Paulsen, I. T., Nelson, K. B., Margatit, I., Read, T. D., Madoff, L. C., Wolf, A. M., Beanan, M. J., Brinkac, L. M., Daugherty, S. C., DeBoy, R. T., Durkin, S., Kolonay, J. F., Umayam, L. A., Madhuri, R., Lewis, M. R., Radune, D., Fedorova, N. B., Scanlan, D., Khouri, H., Mulligan, S., Carty, H. A., Cline, R. T., Gill, J., Scarselli, M., Mora, M., Iacobini, E. T., Bretton, C., Galli, G., Mariani, M., Vegni, F., Malone, D., Rinaudo, D., Rappuoli, R., Telford, J. L., Kasper, D. L., Grandi, G. and Fraser, C. M.

TITLE

Submitted (18-JUL-2002) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

JOURNAL

Location/Qualifiers

1. 21397

/organism="Streptococcus agalactiae 2603V/R"

/mol_type="genomic DNA"

/strain="2603V/R"

/db_xref="taxon:208435"

/complement(155..2605)

/gene="SAG1932"

/complement(155..2605)

/gene="SAG1932"

/note="identified by match to PFM protein family HMM PF02012"

/codon_start=1

/transl_table=11

/product="neuraminidase-related protein"

/protein_id="AA00794.1"

/db_xref="GI:22534975"

/translation="MTLFGKPSITIRLGVITSTYMLGFIILGFLHPTTVEEI

KDINSFISIDKETTVPIDSLDPLTNSIEROLEDSPLSDIYDOKDMVI

VGDSDSOEKLSPPLKSEIATEENKQANLALNOPNRTTIIIPESHSIAOOR

FLIEKGVALLNDSDEPKTARDEGRDRLSDIIPKSDISTSVICNTDIASO

ISLGRKNAIQEHHLTKTRSQDGLSVTEBLAIKGEKEIKYSNLMQKAK

ACGLVCVDNLFEDIKSYQYCHLNQKLPKALISAKTEVVRGDFGRKNDN

FCYRIPSLKTEKGLVAGDERIQAQCMGNIQWIRSEDDGVTKERTVLN

NRPLVLTSGDYSGPSINDMALVODTSKTRFISIDMPEGGVISIANPEY

TOIGOSVYLNNGKKSKVFTIRDKGIYVNRKKTQDVYVITETTKSHSLGDIYK

GKOLLGNYFTTKTSPPRLAKSSVYVMSYSDDGRTWSPDITSLAKOKMKLGI

GGKGIVLKMGPHAGRIIPAYSTWMSKSLRSQSSRLIYSDHGKWTMGAVDNR

ILSNGEKIHSITMDNKEQNTSEVPQLNGDIKLFMRNLTELVATSKDGETWON

LYKRYEVDVAVOLSAIRFEHDKKEYILLVANGGKKRQDGVARLAQVNSGEFKM

LHHHIDGSPAYNSVQOLNDKFGVLYEHRKHQNSFTLANKVFNWSPLSQNTKEQG

TLMEKAAWHLKFEYL"

/complement(2844..4292)

/gene="SAG1933"

/complement(2844..4292)

/gene="SAG1933"

/codon_start=1

/transl_table=11

/product="PTS system, IIC component, putative"

/protein_id="AA00795.1"

/db_xref="GI:22534976"

/translation="MOVFLINVKFEPIIHMGSGVMLIVMTGLAMFGVFSKALE

GGIKALITGAIIGITLGAFSSELOAFVNTGNIISIIDVGNAPLATITWGSPT

LYELLIMLVNIVMTVMKTPDLVDIDPIHMLSTGLIMYAKKNIPILLSVITA

TYAIFVGVKTIINSDMKPTFDDLTIGTPPTSPMSTHNKMNPIINLWLFDFV

PGLDYDPDAALINKAIGFWSKRFITGMLIGLVIGMGVPSFALGDFLGPAG

ACLELFSLIGSWFIAVEBLSQITNFGANGKGRFNIGLDMPTIAGAEIWAQANI

LAPIMLVEAILLSKVGNGILPLAGIILAMGVTPALVITGRILRMITFTGLLPFL

SGTMIAAPATELAKKVGAPAGARAGSLTHTSLBEPMEKIRGYVIGKATTTQSLAI

TLITPATATLGLPMYVAKQKRNAYANQK"

/complement(4332..4637)

/gene="SAG1934"

/complement(4332..4637)

/note="similar to OMN:NTL01SPL345, identified by

sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="PTS system, IIB component, putative"

gene

CDS

/db_xref="GI:22534977"

/translation="MKILACGAVNSHOKDAIPTOLGDRGVNHCDAVAVDIT

EBMNXYFNPIKTKDIFGVNVPVAGSILVIRVSGEYFALBQVKEHNLIS"

/complement(4639..5112)

/gene="SAG1935"

/complement(4639..5112)

/gene="SAG1935"

/codon_start=1

/transl_table=11

/product="PTS system, IIA component, putative"

/protein_id="AA00797.1"

/db_xref="GI:22534978"

/translation="MFAYTODILFIDASHOELEPDLVSKALIKOHVSPNYQAVKER

EREPTGKIDKQDPIQYVAIPHTEOYCVDRIFVYKNSQPIPFKGMIDEECR

VODFFIINRSNQSGLNSLITFFIRKGNLDRHLDGNKKEKINHILIEKGV"

5322..6098

/gene="Iacr-2"

/note="synonym: SAG1936"

5322..6098

/gene="Iacr-2"

/note="identified by match to PFM protein family HMM PF00455"

/codon_start=1

/transl_table=11

/product="lactose phosphotransferase system repressor"

/protein_id="AA00798.1"

/db_xref="GI:22534979"

/translation="MLKREBLQKITEKNINGIVYVIMELBVDSDMTVRDDEL

RAGLIRIHGAQKVNASPTPONKESKSTERYQDTQNTKLELAQAKQFINGETFI

GGTTLEKTLAQDLDFKIRVYVNSLPVNIINQSTLIDLIVGSGYREITGAFVSAT

INSIKLSKSAFVSNQGVPEKSIATYDEGGEIQRIALNNSFEKFLVDSQKFGKYD

FYTFOLDIDPFLVLDHNDVNRKQVSSFKIILNNYM"

/complement(6380..8909)

/gene="SAG1937"

/note="This region contains a gene with one or more

premature stops or frameshifts, and is not the result of a

sequencing artifact; similar to GP:12744742; identified by

sequence similarity; putative; streptococcal histidine

tried family protein; degenerate"

/complement(8936..9859)

/gene="SAG1938"

/complement(8936..9859)

/gene="SAG1938"

/note="SAG1938"

PF01297: The product of this gene was detected by Western

blot analysis. For details on the method see Tettein et

al. 2002."

/codon_start=1

/transl_table=11

/product="adhesion lipoprotein"

/protein_id="AA00799.1"

/db_xref="GI:22534980"

/translation="MKRIKSLIFVLGVVTLICCATKQSQKNGLSVTSFYPVS

ITKAVSGDLNDIKMIRSQSGIHGEPSDDVAIYDADLFYHSHTLMAARLPST

HHSKVSLEAKGKMLDVGHEDEVEAKGVDESTLYPHTNNDVKSSEBQILATO

LAKQOPKAKYQKADQPSDKAMIAKRYKPKFAASKYFVSHTPSYLAKRYGL

TOLGAGSTQGBEAKKLARIQEVKRYKKTITVEGVSKLAQVAASARVYIAS

LSPLAVKRNKNDYDENLETDKVLVSKLNO"

/complement(10114..10557)

/gene="SAG1939"

/complement(10114..10557)

/gene="SAG1939"

/note="identified by match to PFM protein family HMM

PF02580: The product of this gene was detected by Western

blot analysis. For details on the method see Tettein et

al. 2002."

/codon_start=1

/transl_table=11

/product="protein of unknown function TIGR00256"

/protein_id="AA00800.1"

/db_xref="GI:22534981"

/translation="WKIIRVYNQASVISEDVVGSIKGLVLLVGIAPEDTEDIAV

AVRKITSRIFSDDEGKNLISIDIKGSVLSIGFTPLPADTRKGRPAFTGAADPVKA

gene NMFYDIFNOELANHSVETGOGADMOVSLINDGPVITVLDTKNK"
complement (10567 . 12783)
/gene="SAG1940"
CDS complement (10567 . 12783)
/note="identified by match to PPM protein family HM
PF01842"
/codon_start=1
/transl_table=11
/product="GTP pyrophosphokinase family protein"
/protein_id="AA00801.1"
/db_xref="GI:22534982"
/translation="MWKINLTGERVAITSOYSESDVAFKALNYATAHYQAR
KSGEPYIHPHIOVAGLHDLAVTACGLHVEDTEITLDEIETDFGKDYDII
DGATKGVKESHEBOLENNKMLMSDIRITVILKLRDHNHMKLKKDKO
ERISRETEIYAPLARIKISRIKLELSPRYINTEFEFYSIMSEKRESELY
DIYDKIRSTEBOGLYDYGPHITISITRKRDKKRPDQIYDLAIACINETHV
DYVAVGITHELMRPMGRFYDIAAPFANGYQSIHTTVGPKPIEIOIKEMHOV
AEFVAAMAYAKGITSVNQAEQSVGMVQELVELQDESKADPVSXKEDIPE
RIYVTPGAVQELPRESPIIDFAVAIHTQVEKATGAKVGRWVPLAKLTDIVE
IITPNSEFSPRDMKIYKTKARKIKRPFKNODKETSINKRELLVDYQEGYAP
NKYLDKHEIILPRVSVKREALYAAAGPDLSPISFNKLTKEEREERAKAE
ADELINGEITDKRDVLKVSNGVITIQASGLMARIACNPPVPGDILGYITKGR
GVAIHRSDCQNLKSOENYEQRLIDVWDDESKKEYMAEIDIYGLNSGLINDVLT
SNATKLVSTVNAOPTKDMKFNHVSFGISNLAQLTTVDKIKIIPDYVSKRTNG"

gene
/gene="cpdB"
/note="synonym: SAG1941"
CDS 13043 . 15445

Query Match 1.0%; Score 24; DB 1; Length 21397;
Best Local Similarity 100.0%; Pred. No. 0.78;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 485 CGTATACGATGATGTTAT 508
|||||
Db 8327 CGTATACGATGATGTTAT 8304
|||||

RESULT 50
AX602204/c 75248 bp DNA linear PAT 17-FEB-2003
LOCUS
DEFINITION Sequence 133 from Patent WO02092818.
ACCESSION AX602204
VERSION AX602204.1 GI:28402073
KEYWORDS
SOURCE Streptococcus agalactiae
ORGANISM Streptococcus agalactiae
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1
AUTHORS Glaser, P., Rusniok, C., Chevalier, F., Frangeul, L., Lallou, L.,
Zouine, M., Couve, E., Buchrieser, C., Poyart, C., Titeu-Cuot, F., and
Kunst, F.
TITLE Streptococcus agalactiae genome sequence, use for developing
vaccines, diagnostic tools, and for identifying therapeutic targets
JOURNAL Patent: WO 02092818-A 133 21-NOV-2002;
INSTITUT PASTEUR (FR); CENTRE NATIONAL DE LA RECHERCHE
SCIENTIFIQUE (CNRS) (FR)
FEATURES
source Location/Qualifiers
1. 75248
/organism="Streptococcus agalactiae"
/mol_type="genomic DNA"
/db_xref="taxon:1311"
BASE COUNT 24346 a 14447 c 11608 g 24847 t
ORIGIN
Query Match 1.0%; Score 24; DB 6; Length 75248;
Best Local Similarity 100.0%; Pred. No. 0.7;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 485 CGTATACGATGATGTTAT 508
|||||

Db 22369 CGTATACGATGATGTTAT 22346

RESULT 51
SAG766854/c 98050 bp DNA linear BCT 04-NOV-2002
LOCUS
DEFINITION Streptococcus agalactiae NEM316 complete genome, segment 12.
ACCESSION AL766854 AL732656
VERSION AL766854.1 GI:24413438
KEYWORDS
SOURCE Streptococcus agalactiae NEM316
ORGANISM Streptococcus agalactiae NEM316
Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;
Streptococcus.
REFERENCE
1
AUTHORS Glaser, P., Rusniok, C., Chevalier, F., Buchrieser, C., Frangeul, L.,
Zouine, M., Couve, E., Lallou, L., Mesdek, T., Poyart, C.,
Titeu-Cuot, F., and Kunst, F.
TITLE Genome sequence of Streptococcus agalactiae, a pathogen causing
invasive neonatal disease
JOURNAL Mol. Microbiol. 45 (6), 1499-1513 (2002)
MEDLINE 22242508
PUBMED 12354221
2
REFERENCE
1
AUTHORS Glaser, P., Rusniok, C. and Frangeul, L.
TITLE Direct Submission
JOURNAL Submitted (31-MAY-2002) Glaser P., Institut Pasteur, Genomique des
Microorganismes Pathogenes, 25, rue du Docteur Roux, 75724 Paris
Cedex 15, FRANCE. E-mail: pglaser@pasteur.fr Phone: +33 (0)1 45 68
89 96, Fax: +33 (0)1 45 68 87 86
Location/Qualifiers
1. 98050
/organism="Streptococcus agalactiae NEM316"
/mol_type="genomic DNA"
/strain="NEM316"
/serovar="serovar 111"
/db_xref="taxon:211110"
101. 410
/gene="gbs1858"
101. 106
/gene="gbs1858"
117. 410
/gene="gbs1858"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CAD47517.1"
/db_xref="GI:24413440"
/translation="MILINDIALPICEFLVLSAGLLKLEMKHOGYPLERLMTGN
ILTVSLVHLHTLIHINISLFLSITVVLISGKSLISISIVLKRKSKHQV"
complement (405 . 1725)
/gene="gbs1858"
complement (405 . 424)
/gene="gbs1858"
complement (405 . 1725)
/gene="gbs1858"
complement (405 . 1725)
/note="Similar to adenylosuccinate synthase"
/codon_start=1
/transl_table=11
/product="Unknown"
/protein_id="CAD47518.1"
/db_xref="GI:24413440"
/translation="MTSVVVVVGWGMDEGKITTDFLSADAEVIARYOGDNAGHTIV
IDMKRKHLISGIFPEKISIVGNGVYVNPVSKVKEALVHLGEGVTDNLISDA
HVLPHYIKUDLOEDAKDNKIGTTIKGIPAYMDKAAVGRIRADLDREVFAERL
KINLAEKNRLEFKYDSTPLEPDIPEYIYEQQIKQVYTVDSIINDLADGKRVL
FEAGOGVMDLIDGITYFVTSNPNVAGVYIGSGVSPSKINKVVGKATYSRGDGP
FPLELDEVDRIREIGKEXGTTGRPRRVGDSVYMRSRVSGITNSLNSIDL
SGIDTVKICVAVLDGKRIDYVPSLSEQLKCKPIYIELPGWSEDIACRSLLDLPEN
ARVYVRVGEIVGVRISTPSVGRQETNLISVMSNI"
complement (1983 . 3130)
/gene="gbs1860"
complement (1983 . 2012)

gene
terminator

/gene="gbs1860"
 complement(12096..3118)
 /gene="gbs1860"
 /note="similar to transcriptional regulators and to PPS
 enzyme 11"
 /codon_start=1
 /transl_table=1
 /product="Unknown"
 /protein_id="CAD47519.1"
 /db_xref="GI:24413441"
 /translation="NNIIIGTSLILVLAIFTLFNYKAPYGTGKMGALASAAGAPLV
 EAPQSPFGKVLGFQFQLESEGVGANGSLGVAALIVATGPGVAVLGLSISGCG
 IIPGVLAVGFYLYKMERINPGSLDLSITIIIGAPITRVAAKLIPILNSTIIITGG
 DILTGASNSPLIMGLITLGGTIVVAATPLSSMALVAVLGTMPMAIGALSVGSSGF
 MNGVLFHKLKLGSKRDNIAPAVEPLTQADVSNAPIPIYVNPVGAACGLIALMLKL
 VVDITGTTPIPIGFAVMAFVYNNPMIKVILITALGCIILLSLAGFGGIVFKDKVLTKEE
 LQARN"
 /complement(3125..3130)
 /gene="gbs1860"
 3465..4209
 /gene="gbs1861"
 3465..3470
 /gene="gbs1861"
 3480..4154
 /gene="gbs1861"
 /note="similar to unknown proteins"
 /codon_start=1
 /transl_table=1
 /product="Unknown"
 /protein_id="CAD47520.1"
 /db_xref="GI:24413442"
 /translation="MDKEKLDYWKTIITFLHNVLGDNYEIVLHVVDENDIYIGLVNS
 HIGSTIISPLTFALDLIKKKVYEKDFVNNYKIVSPLNKEVRSFTFKKNQDL
 EMLCINIDISAYQNALDIDLVLNANVKLLPSKPSKISLPQEEPEVYVSGIIDI
 ISEIVDSPLNQNHLISOEVKVEISKIHEKGVQLKAGVSKAEVNLISPSYRVL
 KKIETH"
 4188..4209
 /gene="gbs1861"
 4309..6608
 /gene="gbs1862"
 4309..4314
 /gene="gbs1862"
 4324..6576
 /gene="gbs1862"
 /note="similar to gamma-glutamylcysteine synthetase (for
 the N-terminal part) and to cyanophycin synthetase
 (C-terminal part)"
 /codon_start=1
 /transl_table=1
 /product="Unknown"
 /protein_id="CAD47521.1"
 /db_xref="GI:24413443"
 /translation="WITRLQGRSHSLPIIQATFGLEBSLRTHQPPRVAQTPPHK
 TIGSNNAHYHDTQDYSEPOLDITPIADSOEARFLKAIISDVAGRSINHDEYTMPIS
 MPKREBIDIQAOLEDAFEDYRKYLERKTYGKLIQISGHNVLGLQELTSLFEL
 SQADNAIFQNOALYMKLSQNFPLRKMILTYVAGSPVAEEDFLQKLNPPRSIRNSH
 LGVNHKDIRSVTSIKDYVNDLENAVSGQLIAEKERSVPLRGSKYCAKNYEKTI
 TYLERPTDLPNSFPIGTQETVDTVYHLFLALMLIDBSHIIIDODIKARLNDLIML
 SHPLKLPQAFVSDLVAMQSVIOHFNLSFYODLESVRQIQSPELTYAGQLLEML
 IEGSLLETFGQEQQIYHDYMEAPYALKYETMELSTQLLPVIOQGVNFEIDQO
 DOFLKLNHSHIEYKNNMTSKDNYIPLAMANKVYKTLIDKHEFTPGDEFTDR
 KEALYFQIOIDKPIVYKPSKTNFGLGISIFKTSANLASVYKALIDAFETSAALIVEE
 YLEGREYFPIVLEGDCIIVALLRVANVVGDDGIHTISQLYKLNQNPARGDHSRLVEE
 IELGSEVQQLMEQCYTYNLSIPPEGTKEIERNNSISITGSGSIDVTNTMDPTYQQLAA
 EMEKNGMAVCCVDLIFNATQVASKKNAKCIELNPLNMYHTTYQDSGPGSITTP
 RLIALFPEL"
 6588..6608
 /gene="gbs1862"
 6700..7518
 /gene="gbs1863"
 6700..7518
 /gene="gbs1863"
 /note="Similar to unknown proteins"

			/codon_start=1	
			/transl_table=11	
			/product="unknown"	
			/protein_id="CAD47522.1"	
			/db_xref="GI:24413444"	
			/translation="MVKRYKSKNSHNLITLIGIVLASLSDPQSEVIATLLIIGGYC CAVYVDKKRLKRAFTSNORIEALKSDIKETIDDOI RHLEILLKDNRSREYIKLAHQILIP QOLDIREANQLOKALEPNIVKRIITYKANFSEINEQLKLHASPPELEISPOEDDEM IRIAPELKPFYHNIODPFAVLTKKIEADNKAELAAIHQANMKRFTVLGVGRIKOS PKFNNAKERLEQAOLQAIKKRFNDLDDETLQOLNESDWDKPDVSIRMQDBRNKS"	
gene		7504..8821	/gene="gbs1864"	
RBS		7504..7509	/gene="gbs1864"	
CDS		7515..8771	/gene="gbs1864"	
			/note="Similar to other proteins"	
			/codon_start=1	
			/transl_table=11	
			/product="unknown"	
			/protein_id="CAD47523.1"	
			/db_xref="GI:24413445"	
			/translation="MSDTNPFDIDQADNMTKYDKKTLEIIISNOTSOTGOIAFEFKL TPAKSAISEKTPALVDTFPGDONALLDFEQSAVEGNNTTVNHLSQCKIQIOPVDD LLKANRELNGFIARKYKDATPAELEKKNPIIQKLFQSKTSLQEFYDSONIEQMDM MAANVYKOEDTLARNIVASAMLIJEDNTKSIENLYGVAFIESQAEBANASHIQOEI LALDSOTEYOIKNSOARMTREVINTLEOQHTEVSRLYVAWATTPOMRNLVKVSSDM ROKICGMRRNTIPTMKLSINQLGMOOSVSGVTADAIVANNAALOMLAETSKEATIP MBKETRASPTVISIKSVYALAEBSLVYAONGIIAAMDKGRERAQLESVAITSLETINDS VKRIDRKIVEALLNEGSTOKERDES"	
terminator		8799..8821	/gene="gbs1864"	
gene		8911..9786	/gene="gbs1865"	
CDS		8911..9786	/gene="gbs1865"	
			/note="Similar to other proteins"	
			/codon_start=1	
			/transl_table=11	
			/product="unknown"	
			/protein_id="CAD47524.1"	
			/db_xref="GI:24413446"	
			/translation="MDKIIXISITSGSFRAVVLDCTETVRTAOEKHOTLSSITVALGR TLIANQILANQKNSKVYKIVIGDSSFGLIISVADTKGWVGKGIOWTGVIDIKTAGTANG EVLVGPMPNGHFVLTIDVATGOPYSTPLTIGEIGEPAYVLTESSEORPSAVGNLAV LLDDDEDKRVXAGCFMLDYLPQASDEBETSREREKIQEPMPSISLLESNNHSLSLATYY GEDDYKEDSLAFYCDCKSERREALTLTGKEILOAMKDEDGVEITTCFCNQIYY FTBEDLEKIINDSIK"	
gene		9755..10747	/gene="gbs1866"	
RBS		9755..9760	/gene="gbs1866"	
CDS		9770..110747	/gene="gbs1866"	
			/note="Similar to transcription regulator"	
			/codon_start=1	
			/transl_table=11	
			/product="unknown"	
			/protein_id="CAD47525.1"	
			1.0%; Score 24; DB 1; Length 98050;	
			Best Local Similarity 100.0%; Pred.No. 0.69;	
			Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	485	CGCTATACTACAGATGATGGTTAT	S08	
Db	77496	CGCTATACTACAGATGATGGTTAT	77473	
RESULT 52				
LOCUS	AC136885	154101 bp	DNA	linear HTG 26-NOV-2002
DEFINITION	Sus scrofa clone RP44-473j11.	WORKING DRAFT SEQUENCE.		

ACCESSION AC136885
VERSION AC136885.2 GI:25447573
KEYWORDS HTG, HTGS_PHASE2, HTGS_DRAFT.
SOURCE Sus scrofa (pig)
ORGANISM Sus scrofa

REFERENCE
AUTHORS Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE
AUTHORS

TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT
On Nov 26, 2002 this sequence version replaced gi:24819761.
----- Genome Center
Center: NIH Intramural Sequencing Center
Center code: NISC
Web site: <http://www.nisc.nih.gov>
Contact: nisc.zoo@nhgri.nih.gov
----- Project Information
Center project name: dhn
Center clone name: 473J11

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8X average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap, version 0.990119
Consensus quality: 153961 bases at least Q40
Consensus quality: 154051 bases at least Q30
Consensus quality: 154086 bases at least Q20
Insert size: 148000; agarose-fp
Insert size: 154101; sum-of-contigs
Quality coverage: 12.16x in Q20 bases; agarose-fp
Quality coverage: 11.68x in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 1 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

```

FEATURES      *
Source         Location/Qualifiers
              1. 154101: contig of 154101 bp in length.
                1..154101
                  /organism="Sus scrofa"
                  /mol_type="genomic DNA"
                  /db_xref="taxon:9823"
                  /clone="RP4-473j11"
                  /clone_lib="RP4a"
misc_feature   1..154101
               /note="assembly_fragment
               clone_end:17
               vector_side:left
               clone_end:Sp6
               vector_side:right"
misc_feature   1..24598
               /note="clone overlaps with GenBank Accession Number
               AC137063 clone RP4-270E18 (center project name dnm)"
               88502..154101
               /note="clone overlaps with GenBank Accession Number
               AC137536 clone RP4-25AB17 (center project name dbn)"
BASE COUNT    42355 a 33945 c 34061 g 43739 t          1 others
ORIGIN
Query Match   1.0%; Score 23; DB 2; Length 154101;
Best Local Similarity 100.0%; Prod. No. 2,3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy            776 AGTCAGCACTCAAGTAATGA 798
              |||||
Db             107560 AGTCAGCACTCAAGTAATGA 107582

RESULT 53
AC137536      194474 bp DNA linear HTG 18-DEC-2002
LOCUS        Sus scrofa clone RP4-25AB17, WORKING DRAFT SEQUENCE, 7 ordered
DEFINITION   pieces.
ACCESSION    AC137536
VERSION      AC137536.3 GI:27228817
KEYWORDS     HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE       Sus scrofa (pig)
ORGANISM     Sus scrofa
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
REFERENCE    Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
AUTHORS      Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
              Benjamin,B., Blakeley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
              Carajaga,K., Coleman,B., Engle,J., Grantie,S., Guan,X., Gupta,J.,
              Haghighi,P., Han,J., Hansen,N., Ho,S.-I., Idol,J.R., Karlins,E.,
              Latic,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., McDowell,J.,
              Magalies,E.H., Mastello,C., Maesker,B., McDowell,J.,
              Paredis-Dugue,C., Pearson,R., Portnoy,M.E., Prasad,A.,
              Reddix-Dugue,N., Schandler,K., Schueler,M.G., Sison,C.,
              Stenlidop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
              Weherby,K.D., Wiggins,L., Young,A. and Green,E.D.
              NISC Comparative Sequencing Initiative
TITLE        NISC Comparative Sequencing Initiative
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 194474)
AUTHORS      Green,E.D.
TITLE        Direct Submission
JOURNAL      Submitted (22-NOV-2002) NIH Intramural Sequencing Center, 8717
              Grovemont Circle, Gaithersburg, MD 20877, USA
              3 (bases 1 to 194474)
AUTHORS      Green,E.D.
TITLE        Direct Submission
JOURNAL      Submitted (18-DEC-2002) NIH Intramural Sequencing Center, 8717
              Grovemont Circle, Gaithersburg, MD 20877, USA
              On Dec 18, 2002 this sequence version replaced gi:25700111.
COMMENT      ----- Genome Center
              Center: NIH Intramural Sequencing Center
              Center code: NISC
              Web site: http://www.nisc.nih.gov
              Contact: nisc_zoemhgri.nih.gov

```

----- Project Information
 Center project name: dbo
 Center clone name: 254817

The sequence data in this record represents an 'enhanced' version of a Phase 2 submission. Specifically, the indicated order and orientation of each sequence contig has been established using one or more of the following: read-pair data from individual subclones, overlaps with neighboring clones, alignment with available reference sequence (e.g., human), and/or confirmation by PCR testing. In addition, the sequence assembly is based on at least 8x average coverage in Q20 bases and has been reviewed to rule out gross misassemblies, the low-quality ends of sequence contigs have been trimmed away, and each base is associated with a Phrap-derived quality score.

----- Summary Statistics
 Sequencing vector: plasmid; n/a; 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 192852 bases at least Q40
 Consensus quality: 193282 bases at least Q30
 Consensus quality: 193599 bases at least Q20
 Insert size: 175000; agarose-fp
 Insert size: 193874; sum-of-contigs
 Quality coverage: 12.57x in Q20 bases; agarose-fp
 Quality coverage: 11.34x in Q20 bases; sum-of-contigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 7 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the sizes
 * is believed to be correct as given, however the pieces
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 *
 * 1 10381: contig of 10381 bp in length
 * 10382 10481: gap of unknown length
 * 10482 36866: contig of 26385 bp in length
 * 36867 36966: gap of unknown length
 * 36967 100817: contig of 63851 bp in length
 * 100818 100917: gap of unknown length
 * 100918 125399: contig of 24482 bp in length
 * 125400 125499: gap of unknown length
 * 125500 129483: contig of 3984 bp in length
 * 129484 129583: gap of unknown length
 * 129584 160266: contig of 30683 bp in length
 * 160267 160366: gap of unknown length
 * 160367 194474: contig of 34108 bp in length.

FEATURES
 source
 1.194474
 /organism="Sus scrofa"
 /mol_type="genomic DNA"
 /db_xref="taxon:9823"
 /clone="RP4-254817"
 /clone_lib="RP4"
 1..65021
 /note="clone overlaps with GenBank Accession Number AC136885 clone RP4-473J11 (center project name dbn)"
 misc_feature
 1..10381
 /note="assembly_fragment"
 clone_end:SP6
 vector_side:left
 10482..36866
 /note="assembly_fragment"
 36967..100817
 /note="assembly_fragment"
 100918..125399
 /note="assembly_fragment"
 125500..129483
 /note="assembly_fragment"
 129584..160266
 misc_feature

misc_feature
 /note="assembly_fragment"
 160367..194474
 /note="assembly_fragment"
 clone_end:RT
 vector_side:right
 161455..194474
 /note="clone overlaps with GenBank Accession Number AC136887 clone RP4-90B16 (center project name dbp)"

BASE COUNT 52095 a 45169 c 45241 g 51367 t 602 others
 ORIGIN

Query Match 1.0%; Score 23; DB 2; Length 194474;
 Best Local Similarity 100.0%; Pred. No. 2.3;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 776 AGTCAGCAAGTCAGCAAGTATGCA 798
 DB 19120 AGTCAGCAAGTCAGCAAGTATGCA 19142

RESULT 54
 AL732478/c 203434 bp DNA linear ROD 09-JUL-2002
 LOCUS Mouse DNA sequence from clone RP23-42N14 on chromosome 2, complete
 DEFINITION
 ACCESSION AL732478
 VERSION AL732478
 KEYWORDS HTG.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus

REFERENCE
 AUTHORS Direct Submission
 TITLE Submitted (09-JUL-2002) Wellcome Trust Sanger Institute, Hinxton,
 JOURNAL Cambridgehire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
 On Jul 1, 2002 this sequence version replaced gi:21615720.

COMMENT
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: http://www.sanger.ac.uk
 Contact: humquerry@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPep; Information on the WORMPEP database can be found at
 http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-42N14 is from the RPI-23 Mouse Source PAC Library
 constructed by the group of Pieter de Jong.
 For further details see http://www.chori.org/bacpac/home.htm
 VECTOR: pBACe3.6

FEATURES
 source
 1..203434
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /chromosome="2"

```

/clone="RP23-42N14"
/clone.lib="RP21-23"
BASE COUNT      5664 a 44643 c 45404 g 56723 t
ORIGIN
Query Match      1.0%; Score 23; DB 10; Length 203434;
Best Local Similarity 100.0%; Pred. No. 2.3;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      748 TAACACAGACACACACAGCAACA 770
Db      128746 TAACACAGACACACACAGCAACA 128724

RESULT 55
AC115559/c
LOCUS      AC115559      214633 bp      DNA      linear      HTG 18-MAR-2003
DEFINITION Mus musculus clone RP24-65D16, WORKING DRAFT SEQUENCE, 24 unordered
pieces.
ACCESSION      AC115559.4 GI:29029317
VERSION      AC115559.4
KEYWORDS      HTG: HTGS_PHASE1, HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
ORGANISM      Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 214633)
Mus musculus, clone RP24-65D16
Birtten,B., Nusbaum,C. and Lander,E.
Unpublished
2 (bases 1 to 214633)
Birtten,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barua,N., Bassien,V., Bloom,T., Boguslavsky,L.,
Boukhalter,B., Brown,A., Camarata,J., Campolano,A., Chang,J.,
Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collamore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Gilde,S., Gord,S., Goyette,W., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,M., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatae,A., Kelle,C., Lacroque,K., Lamazares,R.,
Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
Maclean,C., Macdonald,P., Major,U., Marquis,N., Matthews,C.,
McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Menus,L.,
Milova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Riebeck,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (22-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 214633)
Birtten,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barua,N., Bassien,V., Bloom,T.,
Boguslavsky,L., Boukhalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collamore,A., Cooke,P., Cook,A., Cooke,P., Corum,B., Dearellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,U., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatae,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
Macdonald,P., Major,U., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Menus,L., Milova,T., Mienga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,

```

```

TITLE
JOURNAL
COMMENT
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (18-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 18, 2003 this sequence version replaced gi:28195861.
All repeats were identified using RepeatMasker:
Smith, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: L24950
Center clone name: 65_D_16
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 206528 bases at least Q40
Consensus quality: 209376 bases at least Q30
Consensus quality: 210668 bases at least Q20
Insert size: 180000; agarose-fp
Insert size: 212333; sum-of-contigs
Quality coverage: 10.6 in Q20 bases; agarose-fp
Quality coverage: 9.0 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 24 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
45098: contig of 45098 bp in length
45099: gap of 100 bp
45198: gap of 100 bp
45199: contig of 751 bp in length
45949: gap of 100 bp
46049: gap of 100 bp
46950: contig of 905 bp in length
46955: gap of 100 bp
47054: gap of 100 bp
47702: contig of 648 bp in length
47703: gap of 100 bp
47802: gap of 100 bp
48563: contig of 761 bp in length
48564: gap of 100 bp
48663: gap of 100 bp
49305: contig of 642 bp in length
49306: gap of 100 bp
49406: gap of 100 bp
49405: contig of 1180 bp in length
50585: gap of 100 bp
50586: gap of 100 bp
50685: gap of 100 bp
51809: contig of 1124 bp in length
51909: gap of 100 bp
51910: gap of 100 bp
53070: contig of 1161 bp in length
53170: gap of 100 bp
53171: gap of 100 bp
53843: contig of 673 bp in length
53844: gap of 100 bp
53943: gap of 100 bp
55624: contig of 1681 bp in length
55944: gap of 100 bp
55724: gap of 100 bp
56906: contig of 1182 bp in length
57006: gap of 100 bp
57007: gap of 100 bp
58348: contig of 1342 bp in length
58448: gap of 100 bp
58449: gap of 100 bp
60304: contig of 1856 bp in length
60404: gap of 100 bp
62280: contig of 1876 bp in length
62380: gap of 100 bp
62381: gap of 100 bp
63184: contig of 1134 bp in length
63515: gap of 100 bp
63615: gap of 100 bp
64733: contig of 1119 bp in length
64833: gap of 100 bp

```

```

* 64834 74899: contig of 10066 bp in length
* 74900 74999: gap of 100 bp
* 75000 86445: contig of 11446 bp in length
* 86545 86545: gap of 100 bp
* 86546 108300: contig of 21755 bp in length
* 108301 108400: gap of 100 bp
* 108401 127566: contig of 19166 bp in length
* 127567 127666: gap of 100 bp
* 182198 182298: contig of 54532 bp in length
* 182199 182298: gap of 100 bp
* 182299 212650: contig of 30351 bp in length
* 212650 212749: gap of 100 bp
* 212750 214633: contig of 1884 bp in length.

```

FEATURES

```

source
1..214633
  /organism="Mus musculus"
  /mol_type="genomic DNA"
  /db_xref="taxon:10090"
  /clone_lib="RP24-65D16"
  /clone_id="RP24 Male Mouse BAC"
  1..45098
  /note="assembly_fragment"
  clone_end:SP6
  vector_side:left"
  45199..45949
  /note="assembly_fragment"
  46050..46954
  /note="assembly_fragment"
  47055..47702
  /note="assembly_fragment"
  47803..48563
  /note="assembly_fragment"
  48664..49305
  /note="assembly_fragment"
  49406..50585
  /note="assembly_fragment"
  50686..51809
  /note="assembly_fragment"
  51910..53070
  /note="assembly_fragment"
  53171..53843
  /note="assembly_fragment"
  53944..55624
  /note="assembly_fragment"
  55725..56906
  /note="assembly_fragment"
  57007..58348
  /note="assembly_fragment"
  58449..60304
  /note="assembly_fragment"
  60405..62280
  /note="assembly_fragment"
  62381..63514
  /note="assembly_fragment"
  63615..64733
  /note="assembly_fragment"
  64834..74899
  /note="assembly_fragment"
  75000..86445
  /note="assembly_fragment"
  86546..108300
  /note="assembly_fragment"
  108401..127566
  /note="assembly_fragment"
  127667..182198
  /note="assembly_fragment"
  182299..212649
  /note="assembly_fragment"
  212750..214633
  /note="assembly_fragment"

```

Query Match 1.0%; Score 23; DB 2; Length 214633;
Best Local Similarity 100.0%; Pred. No. 2.3;

Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2071 AGGCAGAAAGACCACTGTAAG 2093
Db 67890 AGGCAGAAAGACCACTGTAAG 67868

RESULT 56

BT005616

DEFINITION

BT005616 1450 bp mRNA linear PLN 15-MAR-2003
Arabidopsis thaliana clone US0137 putative branched-chain alpha
keto-acid dehydrogenase E1 alpha subunit (At5g09300) mRNA, complete
cds.

ACCESSION

BT005616

VERSION

BT005616.1

KEYWORDS

FLI CDNA

SOURCE

Arabidopsis thaliana

ORGANISM

Arabidopsis thaliana

REFERENCE

1 (bases 1 to 1450)

AUTHORS

Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,
Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,
Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shimn,P.,
Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W.,
Ecker,J.R. and Theologis,A.

TITLE

Arabidopsis Open Reading Frame (ORF) Clones

JOURNAL

Unpublished

REFERENCE

2 (bases 1 to 1450)

AUTHORS

Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,
Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,
Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shimn,P.,
Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W.,
Ecker,J.R. and Theologis,A.

TITLE

Direct Submission

JOURNAL

Submitted (15-MAR-2003) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA

COMMENT

The RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL CDNA : 'RIKEN
Arabidopsis Full-length CDNA') : Seki,M., Narusaka,M., Ishida,J.,
Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai, J.,
Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGECC (SSP) Consortium members constructed and
sequenced the pPUNI (ORF) clones using the RAFL cDNAs: Yamada,K.,
Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M.,
Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C.,
Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y.,
Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M.,
Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shimn,P.,
Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W.,
Ecker,J.R. and Theologis,A.

Yamada,K. (SSP/PGECC) and Seki, M. (RIKEN GSC) contributed equally
to this work. Shinozaki,K. (RIKEN GSC) and Theologis,A. (SSP
/PGECC) contributed equally to this work as PIs.

Annotation based on July 2002 version of the Arabidopsis genome
submitted to Genbank.

FEATURES

```

source
1..1450
  /location/Qualifiers
  /organism="Arabidopsis thaliana"
  /mol_type="mRNA"
  /db_xref="taxon:3702"
  /chromosome="5"
  /clone="mUS0137"
  /note="This clone is in pPUNI 51.
  ecotype: Columbia"
  1..1450
  /gene="At5g09300"

```

gene

CDS

1. 1419
/gene="At5g09300"
/codon_start=1
/evidence=experimental
/product="putative branched-chain alpha keto-acid dehydrogenase E1 alpha subunit"
/protein_id="AA064036.1"
/db_xref="GI:28973423"
/translation="MALHLRSSFSKSTLNLILRNHFGSRSHVTHIRIQLPHDDP LRGSNPSRLCNMAREPILSSQVSEVAVKIYSDMTLQIDNIFIEAQRGLSTFA TAIGBEATNLNLSAALTPQDIVFPQYREPVLLMRGTLQIFANQCFKNSDYKGGQ MPVHGSNKLNFVTSATLQIPNAGAAVSLKMDKACAVYFGDGSSEDFPA ALNIAAWEAVALPFCRNNGWALSTPSDQSRSGVVYKGRAGIRSIKRVNDALAM YSAHTAREMARIEORPLIALTYRGCHSTSDPSYRAGELIEMWNKRPLSRF RTWISNGWMSDIESDIRSIKKEMLEALVAKTERPNQNNFSDYDVPNSNLE QELLVROTINSHPODIYSDVPL"
1420..1450
/gene="At5g09300"
3'UTR
BASE COUNT 406 a 314 c 339 g 391 t
ORIGIN

Query Match 0.9%; Score 22; DB 8; Length 1450;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2285 ACTCTCAATTATGATAC 2306
Db 415 ACTCTCAATTATGATAC 436

RESULT 57
LOCUS AY089131 1547 bp mRNA linear PLN 14-APR-2003
DEFINITION Arabidopsis thaliana clone 3693 mRNA, complete sequence.
ACCESSION AY089131
VERSION AY089131.1 GI:21407905
KEYWORDS FLI CDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1547)
Haas,B.J., Volfovsky,N., Town,C.D., Troupkan,M., Alexandrov,N., Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.
Full-length messenger RNA sequences greatly improve genome annotation
JOURNAL Genome Biol. 3 (6), RESEARCH0029 (2002)
MEDLINE 22088475
PUBMED 12093376
REFERENCE 2 (bases 1 to 1547)
Brover,V., Troupkan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and Feldmann,K.
Full-length cDNA from Arabidopsis thaliana
Unpublished
3 (bases 1 to 1547)
Brover,V., Troupkan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and Feldmann,K.
Direct Submission
Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road, Malibu, CA 90265, USA
COMMENT This clone sequence is one of 5,000 Ceres full-length cDNAs made available to TIGR and Genbank. The following quality assessment of this set was done by comparison with known proteins: two percent of the clones are estimated to be 5'-truncated; less than one percent are 3'-truncated; approximately two percent represent alternative splice variants, including unspliced introns and spliced exons; one percent may contain premature stop codons; five percent may have frame shifts in a coding region. A sequence is considered to be 5'-truncated if it lacks the translation initiation start (ATG). A sequence is considered to be 3'-truncated if it lacks the C-terminal end of the encoded protein. Please note that these cDNA

FEATURES
source
1. 1547
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="3693"
Location/Qualifiers

BASE COUNT 443 a 330 c 347 g 427 t
ORIGIN

Query Match 0.9%; Score 22; DB 8; Length 1547;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2285 ACTCTCAATTATGATAC 2306
Db 447 ACTCTCAATTATGATAC 468

RESULT 58
LOCUS BT004286 1603 bp mRNA linear PLN 14-FEB-2003
DEFINITION Arabidopsis thaliana clone RAFL15-41-A11 (R50137) putative branched-chain alpha keto-acid dehydrogenase E1 alpha subunit (At5g09300) mRNA, complete cds.
ACCESSION BT004286
VERSION BT004286.1 GI:28393750
KEYWORDS FLI CDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 1603)
Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shimn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Arabidopsis Full Length cDNA Clones
Unpublished
2 (bases 1 to 1603)
Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shimn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Direct Submission
Submitted (14-FEB-2003) Plant Gene Expression Center, 800 Buchanan Street, Albany, CA 94710, USA
RIKEN Genomic Sciences Center (GSC) members carried out the collection and clustering of RAFL cDNAs (RAFL cDNA: 'RIKEN Arabidopsis Full-length cDNA'): Seki,M., Narusaka,M., Ishida,J., Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J., Hayashizaki,Y. and Shinozaki,K.
The Salk, Stanford, PEGC (SSP) Consortium members carried out the sequencing and annotation of the RAFL cDNAs: Yamada,K., Chan,M.M., Chang,C.H., Dale,J.M., Hsuan,V.W., Lee,J.M., Onodera,C.S., Quach,H.L., Tang,C.C., Toriumi,M., Wong,C., Wu,H.C., Yu,G., Yuan,S., Carninci,P., Chen,H., Cheuk,R., Hayashizaki,Y., Ishida,J., Jones,T., Kamiya,A., Kawai,J., Kim,C.J., Narusaka,M., Nguyen,M., Palm,C.J., Sakurai,T., Satou,M., Seki,M., Shimn,P., Southwick,A., Tripp,M.G., Wu,T., Shinozaki,K., Davis,R.W., Ecker,J.R. and Theologis,A.
Yamada,K. (SSP/PGEC) and Seki,M. (RIKEN GSC) contributed equally to

this work. Shinozaki, K. (RIKEN GSC) and Theologis, A. (SSP/PGEC) contributed equally to this work as PIs.

Annotation based on July 2002 version of the Arabidopsis genome submitted to Genbank.

FEATURES

source

Location/Qualifiers

1..1603

/organism="Arabidopsis thaliana"

/mol_type="mRNA"

/db_xref="taxon:3702"

/chromosome="5"

/clone="RAF15-41-A11 (R50137)"

/note="This clone is in a modified pBluescript vector2 (lambda PS) as a BamHI/XhoI insert."

ecotype: Columbia"

1..1603

/gene="At5g09300"

1..66

/gene="At5g09300"

67..1485

/gene="At5g09300"

/codon_start=1

/evidence="experimental"

/product="putative branched-chain alpha keto-acid dehydrogenase E1 alpha subunit"

/protein_id="AA042286.1"

/db_xref="GI:28393751"

/translation="MALHRSFSSKSTLNLIRHNGSRSHVTRHRIOLPHDP

LRGSONPISRICNTMAEPTLSPFOHEYANNHQMDFPGCVATPEIOTFSEDEK

RVCYRLVDNGQLITNSOFVQSEVAVKITSDDVTQIMNITEYERQORLSFVA

TAIGEAINTASAAALTPQDVIFPQYREGVILMFGFTLOEPANOCFENKSGYGRQ

MPVHGSKLNFATVATQLPNPAVAGSLKMDKACAVTFGCGSGGPFHA

ALNIAVMEAPVLFICRNNGMAISPTSDQFSDGVVKGRAVGRSIRVDNDALAM

YSATVAREMAIREORPLIEALTYRGHSHSDSTYRSAGEIEMNNKAPNLSRP

RTWISNEMMSKTESDARSIRKEMLEALRAEKTETKPNLQNMESDYVDPPSLRE

QELVROITINSFPDYPSPDVL"

1486..1603

/gene="At5g09300"

1587

/gene="At5g09300"

/note="artifact within poly A tail"

BASE COUNT

471 a 341 c 349 g 442 t

ORIGIN

Query Match Best Local Similarity 0.9%; Score 22; DB 8; Length 1603;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2285 ACTCTCAATTATGATACCA 2306

Db 481 ACTCTCAATTATGATACCA 502

RESULT 59

AC015298

LOCUS Drosophila melanogaster, *** SEQUENCING IN PROGRESS ***

AC015298

AC015298

AC015298

AC015298

AC015298

AC015298

AC015298

AC015298

AC015298

AC015298

AC015298

AC015298

AC015298

AC015298

AC015298

AC015298

AC015298

AC015298

AC015298

* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES

source

Location/Qualifiers

1..54419

/organism="Drosophila melanogaster"

/mol_type="genomic DNA"

/db_xref="taxon:7227"

BASE COUNT

15977 a 15728 c 10928 g 14786 t

ORIGIN

Query Match

Best Local Similarity 0.9%; Score 22; DB 2; Length 54419;

Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 755 AGCACAACACGACACCTACCA 776

Db 28695 AGCACAACACGACACCTACCA 28716

RESULT 60

ATT5E8/c

LOCUS

DEFINITION Arabidopsis thaliana DNA chromosome 5, BAC clone T5E8 (ESSA

project).

ACCESSION

AL391712

KEYWORDS

SOURCE

ORGANISM

Arabidopsis thaliana (thale cress)

Arabidopsis thaliana

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;

rosids; eurooids II; Brassicales; Brassicaceae; Arabidopsis.

1 (bases 1 to 103534)

Bevan, M., Murphy, G., Ridley, P., Hudson, S., Bancroft, I., Mewes, H. W.,

Rudd, S., Lemcke, K. and Mayer, K. F. X.

Unpublished

2 (bases 1 to 103534)

EU Arabidopsis sequencing project.

Direct Submission

Submitted (28-AUG-2000) MIPS, at the Max-Planck-Institut fuer

Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail:

lemcke@mips.biochem.mpg.de, mayer@mips.biochem.mpg.de Project

Coordinator: Mike Bevan, Molecular Genetics Department, Cambridge

Laboratory, John Innes Centre, Colney Lane, NR4 7UJ Norwich, UK,

E-mail: michael.bevan@bbsrc.ac.uk

Information on performance of analysis and a more detailed

annotation of this entry and other sequences of chromosomes 3, 4

and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.

FEATURES

source

Location/Qualifiers

1..103534

/organism="Arabidopsis thaliana"

/mol_type="genomic DNA"

/variety="Columbia"

/db_xref="taxon:3702"

/chromosome="5"

3395..6144

/gene="T5E8.10"

/complement[Join(3395..3451,3525..3633,4011..4306,

4412..4685,4910..5226,5275..5397,5442..5567,5635..6144)]

/gene="T5E8.10"

complement[Join(3395..3451,3525..3633,4011..4306,

4412..4685,4910..5226,5275..5397,5442..5567,5635..6144)]

/gene="T5E8.10"

/note="similarity to GCF2 fusion protein - Homo sapiens,

EMBL:AB026911"

/codon_start=1

/product="putative protein"

/protein_id="CAC05447.1"

/db_xref="GI:9955508"

/translation="WGTRKQPNRYRRNCDEIDEDATAAKPSSSDLYPAKPSDDLY

PKWRKRLPAITKDDDAKGRDAKSSWLLDLSGDEFTYQEQQRLEKRSQDERPM

EIVNGERVRISIGMDGLSYFVCEPQTADDKMALDRICNAYDERVSRMPQEPQM

/gene="T5E8_30"
 /number=4
 12990..13089

intron

Query Match 0.9%; Score 22; DB 8; Length 103534;
 Best Local Similarity 100.0%; Pred. No. 8.3;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2285 ACTCTTCAATTATGATACCA 2306
 |||||||
 Db 26781 ACTCTTCAATTATGATACCA 26760

RESULT 61
 AC123445/c
 LOCUS
 DEFINITION
 AC123445
 Rattus norvegicus clone CH230-403G9, *** SEQUENCING IN PROGRESS
 **, 3 unordered pieces.
 AC123445
 AC123445.4 GI:25137917
 HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
 Rattus norvegicus (Norway rat)
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (baes 1 to 168601)
 Munry,D,Marie., Metzker,M, Lee., Abramzon,S., Adams,C., Alder,J.,
 Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angiano,D.,
 Anyalbechti,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
 Biswalto,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
 Bryant,N., Buhay,C., Burck,P., Burrell,K., Calderon,E.,
 Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Center,A.,
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
 Devila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
 Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Diya,K.,
 Diaper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
 Fraser,C.M., Gabisi,A., Ganti,R., Garcia,A., Garner,T., Garza,M.,
 Georgeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,M., Guevara,W.,
 Gunaratne,P., Haaland,W., Hamli,C., Hamilton,C., Hamilton,K.,
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
 Hernandez,J., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M.,
 Hollins,B., Howells,S., Huljk,S., Hume,J., Idlebird,D., Jackson,A.,
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
 Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
 Kowls,C., Kieft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
 Liu,Y., Liu,Y., Liu,Y., London,P., Longacre,S., Lopez,J.,
 Lorenshewa,L., Louissege,H., Lozado,R.J., Lu,X., Ma,J.,
 Maheshwari,M., Mahindratne,M., Mahmoud,M., Malloy,K., Mangum,A.,
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
 Mawhinley,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
 Milosavljevic,A., Miner,G., Mnja,E., Montemayor,J., Moore,S.,
 Morgan,M., Morris,K., Morris,S., Mundasa,M., Murphy,M., Nair,L.,
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
 Nwokilemeh,O., Okunou,G., Olampunsaogon,A., Pal,S., Parks,K.,
 Paternak,S., Paul,H., Perez,A., Perez,L., Pfamkoch,C.,
 Plopper,F., Poldexler,A., Popovic,D., Prims,E., Pu,L., L.,
 Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Reiter,M.A., Reigh,R.,
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
 Sanders,W., Savery,G., Scherer,S., Scott,G., Shatman,S., Shen,H.,
 Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
 Steimle,M., Strong,R., Sutton,A., Syatke,A., Tabot,P., Taylor,C.,
 Taylor,T., Thomas,N., Thomas,S., Tinney,A., Trejos,Z., Usmani,K.,
 Valas,R., Vera,V., Villasana,D., Waliron,L., Walker,B., Wang,J.,
 Wang,Q., Wang,S., Warren,R., Wiczek,R., Woden,H., Worley,K.,
 Williams,G., Willson,R., Wiczek,R., Woden,H., Worley,K.,
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,

```

Yu.F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Wenstoeck, G. and Gibbs, R.A.
TITLE
JOURNAL
Direct Submission
Unpublished
2 (bases 1 to 168601)
REFERENCE
AUTHORS
Moley, K.C.
TITLE
Direct Submission
Submitted (29-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 168601)
REFERENCE
AUTHORS
Rat Genome Sequencing Consortium.
TITLE
Direct Submission
Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:23907751.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu

----- Project Information
Center project name: GXID
Center clone name: CH230-403G9

----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 159374 bases at least Q40
Consensus quality: 160849 bases at least Q30
Consensus quality: 161918 bases at least Q20
Estimated insert size: 162538; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 165292: contig of 165292 bp in length
* 165293 165293: gap of unknown length
* 165383 166779: contig of 1387 bp in length
* 166780 166879: gap of unknown length
* 166880 168601: contig of 1722 bp in length.
*
* Location/Qualifiers
* 1. 168601
* /organism="Rattus norvegicus"
* /mol_type="genomic DNA"
* /db_xref="taxon:10116"
* /clone="CH230-403G9"
* 1. 1738
* /note="wgs_end_extension
* clone_end:5p6"
* /note="wgs_end_extension
* clone_end:5p6"

```

```

misc_feature      7303..8829
                  /note="wgs_end_extension
                  clone_end:Sp6"
misc_feature      9046..10005
                  /note="Clone_boundary
                  clone_end:Sp6
                  site:
misc_feature      end_sequence:BZ274038"
                  complement(160964..161915)
                  /note="Clone_boundary
                  clone_end:T7
                  site:
misc_feature      end_sequence:BZ274036"
                  163560..165292
                  /note="wgs_end_extension
                  clone_end:T7"
BASE COUNT      45049 a 35319 c 35438 g 46783 t 6012 others
ORIGIN
Query Match      0.9%; Score 22; DB 2; Length 168601;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      1675 AGATGACCTTCTGATAAGAA 1696
        |||||||
Db      112000 AGATGACCTTCTGATAAGAA 111979

RESULT 62
AC007811/c
LOCUS
DEFINITION      AC007811      171569 bp      DNA      linear      INV 15-FEB-2001
Drosophila melanogaster, chromosome 3R, region 90C-90D, BAC clone
ACCESSION      AC007811
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
Ceoliker,S.E., Adams,M.D., Kronmiller,B., Tyler,D., Wan,K.H.,
Holt,R.A., Evans,C.A., Gockayne,J.D., Amanatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A.,
Carlson,J.W., Center,A., Champe,M., Davenport,L.B., Dietz,S.M.,
Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Farfan,D.,
Fertiera,S., Frisze,E., Galle,R.F., Garg,N.S., George,R.A.,
Gonzalez,M., Houck,J., Hoskins,R.A., Hostrin,D., Howland,T.J.,
Ibbagang,C., Jalali,M., Kruse,D., Li,P., Matzel,B., Moshrefi,A.,
McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Munoz,J.,
Paclet,J., Paragas,V., Park,S., Patel,S., Pfeiffer,B.,
Phouananaong,S., Piltman,G.S., Puri,V., Richards,S., Scheeler,F.,
Stepleton,M., Strong,R., Svirskas,R., Tector,C., Williams,S.M.,
Zaveri,J.S., Smith,H.O., Rubin,G.M. and Venter,J.C.
Sequencing of Drosophila chromosome 3R, region 90C-90D
Unpublished
2 (bases 1 to 171569)
Ceoliker,S.E., Abhayani,A., Arcaina,T.T., Baxter,E., Blazej,R.G.,
Buenhoff,C., Champe,M., Chavez,C., Chew,M., Ciesiolka,L.,
Doyle,C.M., Farfan,D.E., Galle,R., George,R.A., Harris,N.L.,
Hoskins,R.A., Houston,K.A., Hummasti,S.R., Katta,K., Kearney,L.,
Kim,E., Lee,B., Lewis,S., Li,P., Lomtan,M.A., Mazda,P.,
Moshrefi,A.R., Moshrefi,M., Nixon,K., Paclet,J.M., Park,S.,
Pfeiffer,B., Poon,L., Seguirra,A., Sethi,H., Shih,E.,
Svirskas,R.R., Wan,K.H., Weinburg,T., Zhang,R., Zieran,L.L. and
Rubin,G.M.
Direct Submission
Submitted (14-JUN-1999) Drosophila Genome Center, Lawrence Berkeley
Laboratory, MS 64-121, Berkeley, CA 94720, USA
On Feb 15, 2001 this sequence version replaced gi:5822638.
Sequence submitted by:
Berkeley Drosophila Genome Project

```

Lawrence Berkeley National Laboratory, MS 64-121
 Berkeley, CA 94720
 For further information about this sequence, including its location
 and relationship to other sequences, please visit our sequence
 archive web site (<http://www.fruitfly.org/sequence/>) or send email
 to bdg@fruitfly.berkeley.edu.
 location/Qualifiers

FEATURES
 SOURCE

1. 171569
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /strain="y; cn bw sp"
 /db_xref="taxon:7227"
 /chromosome="3R"
 /map="90C-90D"
 /clone="BAC BACR45M04 (D721)"
 /clone_id="RPCT1-98 (Roswell Park Cancer Institute
 Drosophila melanogaster BAC library, partial ECORI in
 PBAC3.6)"

BASE COUNT
 47106 a 36072 c 38013 g 50378 t

ORIGIN

Query Match 0.9%; Score 22; DB 3; Length 171569;
 Best Local Similarity 100.0%; Pstd. No. 8;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 755 AGCAACAACAGCACTACTACA 776
 ||||||||||||||||||||
 Db 147757 AGCAACAACAGCACTACTACA 147736

RESULT 63
 HSBA192P3/c 187517 bp DNA linear PRI 04-MAR-2003
 LOCUS Human DNA sequence from clone RP11-192P3 on chromosome 10, complete
 DEFINITION sequence.
 AL117340 AL117340.3 GI:6002147
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 187517)
 Bird.C.
 Direct Submission
 Submitted (04-MAR-2003) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humquerry@sanger.ac.uk
 Clone requests: clonerequest@sanger.ac.uk
 On Sep 30, 1999 this sequence version replaced gi:5931851.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 The following abbreviations are used to associate primary accession
 numbers given in the feature table with their source databases:
 Em; EMBL; Sw; SWISSPROT; Tr; TREMBL; Wp; WORMPEP; Information
 on the WORMPEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep
 This sequence was generated from part of bacterial clone contigs of human
 chromosome 10, constructed by the Sanger Centre Chromosome 10
 Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr10>
 ----- Genome Center
 Center: Wellcome Trust Sanger Institute
 Center code: SC
 Web site: <http://www.sanger.ac.uk>
 Contact: humquerry@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest, except on the rare occasion of the clone being a YAC. RP11-152P3 is from the library RP11-11.1 constructed by the group of Pieper de Jong. For further details see <http://www.chori.org/bacpac/home.htm>
VECTOR: pBAC3.6.

FEATURES

source

1..187517
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-152P3"
/clone_1fb="RP11-11.1"
BASE COUNT 57499 a 33087 c 35463 g 61468 t
ORIGIN

Query Match 0.9%; Score 22; DB 9; Length 187517;
Best Local Similarity 100.0%; Pred. No. 7.9;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1905 ATCATTACCATATATTAATT 1926
|||||
Db 140645 ATCATTACCATATATTAATT 140624

RESULT 64
AF225898 198084 bp DNA linear PRI 23-FEB-2000
LOCUS Homo sapiens BAC clone 13d21, complete sequence.
DEFINITION
ACCESSION AF225898
VERSION AF225898.1 GI:7021527

KEYWORDS

SOURCE

ORGANISM Homo sapiens (human)
Eukaryota; Eukarya; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 198084)
Fitzpatrick,E.S., Hammond,H.A., DeAngelis,D.M., Soderman,A.R.,
Wright,J.L., Liu,X., Diane,L., McGowan,J., Ziegler,S., Hess,J.F.,
Todd,J., Caskey,C.T. and Metzker,M.L.
Homo sapiens Chromosome 10 BAC clone b13-d-21

TITLE

JOURNAL

2 (bases 1 to 198084)
Fitzpatrick,E.S.
Direct Submission
Submitted (18-JAN-2000) Department of Human Genetics, Merck & Co.
Inc, SumneyTown Pike, West Point, PA 19486, USA

JOURNAL

TITLE

JOURNAL

Location/Qualifiers

source

1..198084
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="BAC 13d21"
/complement(3383..3451)
/rpt_family="Alu"
complement(5194..5444)
/rpt_family="Alu"
complement(5480..5608)
/rpt_family="MER2"
complement(6514..7048)
/rpt_family="L1"
7337..7411
/rpt_family="Alu"
complement(7901..8013)
/rpt_family="Alu"
12020..12266
/standard_name="L01011"
repeat_region 18754..19038

/rpt_family="Alu"
20222..20380
complement(120629..20808)
/rpt_family="MER3"
complement(120987..21277)
/rpt_family="Alu"
complement(122542..22828)
/rpt_family="Alu"
complement(126710..26947)
/rpt_family="MER33"
28161..28383
/standard_name="216795"
complement(129851..30141)
/rpt_family="Alu"
complement(34033..34153)
/rpt_family="Alu"
complement(34206..34245)
/rpt_family="L1"
complement(36005..36285)
/rpt_family="Alu"
complement(41998..42066)
/rpt_family="MER45"
complement(42726..48097)
/rpt_family="L1"
44643..46379
/rpt_family="L1"
48377..48669
/rpt_family="Alu"
complement(49516..49664)
/rpt_family="L1"
complement(55827..56017)
/rpt_family="MER46"
complement(59936..60051)
/rpt_family="L1"
64590..64886
/rpt_family="Alu"
complement(64910..66418)
/rpt_family="L1"
complement(66758..67045)
/rpt_family="Alu"
69176..69312
/rpt_family="MER3"
69332..69591
/rpt_family="Alu"
71973..72437
/rpt_family="MLT1"
75719..75836
/rpt_family="Alu"
77306..79590
/rpt_family="L1"
complement(80429..86448)
/rpt_family="L1"
complement(84988..85206)
/rpt_family="MER25"
complement(86922..87050)
/rpt_family="MER"
90653..93427
/rpt_family="L1"
94384..198084
/note="Region: ACO05877 Homo sapiens chromosome 10 clone
CT9878X-1019018 map"
95210..95444
/standard_name="223391"
97612..97702
/rpt_family="MER"
join(98048..98249,132795..132849,139367..139528,
147695..147897,151625..151730,157139..158949,
160946..161130,163685..165932)
/product="Two-handed zinc finger protein ZEB"
/note="Corresponds to mRNA sequence deposited in GenBank
Accession Number U19969"
join(98048..98249,132790..132849,139367..139528,

147695..147897,151625..151730,157139..158949,
 160946..161126,163685..164819)
 /note="zinc finger homeodomain protein"
 /product="corresponds to mRNA sequence deposited in Genbank
 Accession Number U12170"
 join(98048..98253,132795..132849,139367..139528,
 147695..147897,151625..151727,157139..158949,
 160946..161126,163685..164393)
 /product="transcription factor AREB6"
 /note="corresponds to mRNA sequence deposited in Genbank
 Accession Number D15050"
 101246..101381
 /rpt_family="Alu"
 repeat_region complement(102675..102795)
 /rpt_family="Alu"
 repeat_region 105618..105688
 /rpt_family="MER25"
 repeat_region 105841..106220
 /rpt_family="MER25"
 repeat_region 106230..112234
 /rpt_family="L1"
 repeat_region 107460..107678
 /rpt_family="MER25"
 repeat_region 112242..112312
 /rpt_family="MER25"
 repeat_region 113105..115797
 /rpt_family="L1"
 repeat_region 118138..118492
 /rpt_family="MER42"
 repeat_region 118577..118621
 /rpt_family="Alu"
 repeat_region 118730..119132
 /rpt_family="MER42"
 repeat_region complement(120797..121292)
 /rpt_family="MT1"
 repeat_region complement(123425..123579)
 /rpt_family="L1"
 repeat_region complement(123597..123668)
 /rpt_family="L1"
 repeat_region 136623..136794
 /rpt_family="Alu"
 repeat_region 138901..138953
 /rpt_family="MER44C"
 repeat_region complement(140450..142773)
 /rpt_family="L1"
 repeat_region complement(144024..144211)
 /rpt_family="L1"
 repeat_region complement(144531..144748)
 /rpt_family="MER25"
 repeat_region complement(154864..154952)
 /rpt_family="MIR"
 repeat_region complement(161217..161511)
 /rpt_family="Alu"
 STS 163682..165932
 /note="human STS SHGC-31480"
 STS 16588..166812
 /note="human STS SHGC-15789"
 repeat_region complement(167450..167734)
 /rpt_family="Alu"
 repeat_region 171106..171210
 /rpt_family="L1"
 repeat_region 171783..172068
 /rpt_family="Alu"
 repeat_region complement(173315..175503)
 /rpt_family="Alu"
 repeat_region complement(175843..176175)
 /rpt_family="L1"
 repeat_region complement(176497..176789)
 /rpt_family="Alu"
 repeat_region 179341..179505
 /rpt_family="MER4"
 repeat_region complement(179601..179677)
 /rpt_family="MIR"

repeat_region 179845..180184
 /rpt_family="MER21"
 repeat_region 181513..181648
 /rpt_family="THE1"
 repeat_region 181649..182174
 /rpt_family="WSTAR"
 repeat_region complement(182278..182604)
 /rpt_family="THE1"
 repeat_region complement(182651..184166)
 /rpt_family="WSTAR"
 repeat_region complement(184171..184684)
 /rpt_family="THE1"

Query Match 0.9%; Score 22; DB 9; Length 198084;
 Best Local Similarity 100.0%; Pred. No. 7.9;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1905 ATCATACCATATATTAATT 1926
 Db 23057 ATCATACCATATTAATT 23036

RESULT 65
 AE003719/c 212941 bp DNA linear INV 14-FEB-2003
 LOCUS Drosophila melanogaster chromosome 3R, section 57 of 118 of the
 DEFINITION complete sequence.
 ACCESSION AE003719 AE002708 AE014297
 VERSION AE003719.3 GI:23171576
 KEYWORDS
 SOURCE
 ORGANISM Drosophila melanogaster (fruit fly)
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 212941)
 Adams,M.D., Celnikier,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
 Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galie,R.F.,
 George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
 Sutton,G.G., Wortman,J.R., Vandal,M.D., Zhang,Q., Chen,L.X.,
 Brandon,R.C., Rogers,Y.H., Blazer,V.G., Nelson,C.R., Gaber,G.L.,
 Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Neimann,C.R., Baldwin,D.,
 Abell,J.F., Agbayani,A., An,H.J., Andrews-Pfannkoch,C., Baldwin,D.,
 Ballew,R.M., Basu,A., Baxendale,J., Bayraktaroglu,L., Beasley,E.M.,
 Beeson,K.Y., Benos,P.V., Bernan,B.P., Bhandari,D., Bolshakov,S.,
 Bokora,D., Botchan,M.R., Bouck,J., Brockstein,P., Brotlier,P.,
 Burris,K.C., Busam,D.A., Butler,H., Cadieu,E., Center,A.,
 Chandrasekhar,I., Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B.,
 Davies,P., de Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I.,
 Dietz,S.M., Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S.,
 Dunkov,B.C., Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C.,
 Ferrieres,S., Fleischmann,W., Foster,C., Gabriellian,A.E., Garg,N.S.,
 Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,U.H., Gu,Z.,
 Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
 Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
 Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
 Kramson,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
 Kravitz,S., Kulp,D., Lai,Z., Laoko,P., Lei,Y., Levitsky,A.A.,
 Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
 McLeod,M.P., McPherson,D., Merkulov,G., Mishina,N.V., Mobarry,C.,
 Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
 Muray,D.M., Nelson,D.L., Nelson,D.R., Nelson,K.A., Nixon,K.,
 Nusskern,D.R., Pacle,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
 Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
 Sanders,R.D., Scheeler,F., Shen,H., Spier,B.C., Siden-Kiamos,I.,
 Simpson,M., Skupski,M.P., Smith,T., Spate,E., Spradling,A.C.,
 Stapleton,M., Strong,R., Sun,E., Svirskaas,R., Tector,C., Turner,R.,
 Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassenaar,D.A.,
 Weinstock,G.M., Weissbach,J., Williams,S.M., Woodgett,
 Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yen,R.F.,
 Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
 Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
 Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.

REFERENCE
 AUTHORS

TITLE The genome sequence of *Drosophila melanogaster*
JOURNAL Science 287 (5461), 2185-2195 (2000)
MDLINE 20196006
PMID 10731132
REFERENCE 2 (bases 1 to 212941)
AUTHORS Evans,C.A., Gockyne,J.D., Amanatides,P.G., Brandon,R.C., Rogers,Y., Banzon,J., An,H., Baldwin,D., Banzon,J., Beeson,K.Y., Busam,D.A., Carlson,J.W., Center,A., Champagne,M., Davenport,L.B., Dietz,S.M., Dodson,K., Dorsett,V., Doup,L.E., Doyle,C., Dresnek,D., Fartan,D., Ferreira,S., Frisze,E., Galle,R.F., Garg,N.S., George,R.A., Gonzalez,M., Houck,J., Hoskins,R.A., Hostin,D., Howard,T.J., Ileguam,C., Jalali,M., Kruse,D., Li,P., Mattei,B., Moshrefi,A., McIntosh,T.C., Moy,M., Murphy,B., Nelson,C., Nelson,K.A., Nuno,J., Paclet,J., Paragas,V., Park,S., Pacel,S., Pfeiffer,B., Phouanavong,S., Piltman,G.S., Puri,V., Richards,S., Scheeler,F., Stapleton,M., Strong,R., Svitskas,R., Tector,C., Tyler,D., Williams,S.M., Zaveri,J.S., Smith,H.O., Venter,J.C. and Rubin,G.M.
TITLE Sequencing of *Drosophila melanogaster* genome
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 212941)
AUTHORS Misra,S., Crosby,M.A., Matthews,B.B., Bayraktaroglu,L., Campbell,K., Hradecky,P., Huang,Y., Kaminker,J.S., Prochuk,S.E., Smith,C.D., Tupy,J.L., Bergman,C.M., Bertan,B.P., Carlson,J.W., Clunker,S.E., Clump,M.E., Drysdale,R.A., Emmert,D., Frisze,E., de Grey,A.D.N.U., Harris,N.L., Krommiller,B., Marshall,B., Mileyburn,G.H., Richter,J., Russo,S., Searle,S.M.J., Smith,E., Shu,S., Smutnicki,J., Whitfield,E.J., Ashburner,M., Gelbart,W.M., Rubin,G.M., Munnali,C.J. and Lewis,S.E.
TITLE Annotation of *Drosophila melanogaster* genome
JOURNAL Unpublished
REFERENCE 4 (bases 1 to 212941)
AUTHORS Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,J.C.J.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive, Rockville, MD 20850, USA
REFERENCE 5 (bases 1 to 212941)
AUTHORS FlyBase
CONSTRM Direct Submission
TITLE Submitted (06-SEP-2002) University of California Berkeley, 539 Life Sciences Addition, Berkeley, CA 94720, USA
JOURNAL 6 (bases 1 to 212941)
REFERENCE FlyBase
AUTHORS Direct Submission
CONSTRM Submitted (13-FEB-2003) University of California Berkeley, 539 Life Sciences Addition, Berkeley, CA 94720, USA
TITLE Sequence update by Submitter
REMARK On Sep 10, 2002 this sequence version replaced gi:10726587.
COMMENT location/Qualifiers
FEATURES 1. 212941
SOURCE /organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/chromosome="3R"
/note="genotype: Y[1]; cn[1] bw[1] sp[1]; Rh6[1]"
583..4066
/gene="pxt"
/locus_tag="CG7660"
/note="Last curated on Mon Aug 26 10:55:58 PDT 2002"
/map="90C2-90C2"
/db_xref="FLYBASE:FBgn0038538"
join(583..845,1140..1927,2060..2552,2615..2942,3004..3178,3241..3399,3454..3773,3829..4066)
/gene="pxt"
/locus_tag="CG7660"
/product="CG7660-PA"
/db_xref="FLYBASE:FBgn0038538"
join(583..845,1140..1927,2060..2552,2615..2942,3004..3178,3241..3399,3454..3773,3829..4066)
/gene="pxt"
/locus_tag="CG7660"

CDS /product="CG7660-RB"
/db_xref="FLYBASE:FBgn0038538"
join(680..845,1140..1927,2060..2552,2615..2942,3004..3178,3241..3399,3454..3773,3829..3895)
/gene="pxt"
/locus_tag="CG7660"
/note="pxt gene product from transcript CG7660-RA"
/codon_start=1
/product="CG7660-PA"
/protein_id="AAFS5458.2"
/db_xref="GI:23171577"
/db_xref="FLYBASE:FBgn0038538"
/translation="MSRLFTLLLLITVQLSEIQAAFSVRNRFDEVPDLQTPAPLA
TSTSSKPEKATSGLLKCLPCSDGIRCVPOIQPAHVMESEKFOICDLPGKRG
YCEGTGQHTAPKPEKSPKERSGFTLLSPAVLEARNNEHLMHGAOVLPAKRG
DPAHGLVHSTAKDLNFIANSALQVWTTQIFGKEQVPEDFITNNVPIKFTT
PLAHGCPYPVCNIRSVYRMDGTCKNPEQSLMAAQQPMERMLPAYEDGWTW
RAHSDGTPFLGARKIRSLSDVDRHPKTNLMWFGVLAHDSISIRIREDSS
LVQCCSPGKVALSPQOSHFAICMPIHVEPDEFSAQCLNFRSLVSPSPCOIS
YKQILTKVTHFVDASPVYSSDEASRLAFRGRLRMNDFGDLPLTNDKACPS
EEAGKSCFHSQDGRNTQISLITQILAEHNVAGALHELNDSASDELTFQGAR
VIAVMQHTVNEPFLITIIPOOMKRFPLPHQSGSYDVVNVNPAITNFEQAAY
GHSYVDGKFOIRQHGRIDEVYNIPTDFNPNRKRKEFTDDMLRTIYSPMOVDS
ISQGISRFLFERGNDPFGDILANINIQRGQGLNSYDYLELWKAFLHSFEOQPEI
AOKLSRVYRTPDDIDLWVGILAEVAGVGAVGVPAEIIADQFARFGQGRYYEYDN
GINQAFNPDLQDEIRKVTALRLCDNSDRITLQAVLAAFAVRADHPGNMIGCDPN
LPSVNLBAMRA"
complement(553..6153)
/locus_tag="CG12334"
/note="Last curated on Wed Jan 23 07:52:15 PST 2002"
/map="90C2-90C2"
/db_xref="FLYBASE:FBgn0038539"
complement(553..6153)
/locus_tag="CG12334"
/product="CG12334-PA"
/db_xref="FLYBASE:FBgn0038539"
complement(567..6019)
/locus_tag="CG12334"
/note="CG12334 gene product"
/codon_start=1
/product="CG12334-PA"
/protein_id="AAFS5459.1"
/db_xref="GI:7300298"
/db_xref="FLYBASE:FBgn0038539"
/translation="MDNNVOYKXDSFDRKNEGDKIRRYKVPDEVPVIVKAPKTRVA
ELDKKYIVPADLIVGQFYFLIRKIRNLRRDDALFFVNVNVIPTTSATMGALVQEHED
KDYFLIYSTDENYVGRQ"
6635..14339

gene
CDS
gene

Query Match 0.9%; Score 22; DB 2; Length 221952;
 Best Local Similarity 100.0%; Pred. No. 7.8;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 318 TTATCAAGTGAATGGAATA 339
 Db 49170 TTATCAAGTGAATGGAATA 49191

RESULT 67
 AC0989902.C
 LOCUS AC0989902 221952 bp DNA linear HTG 10-MAY-2003
 DEFINITION Rattus norvegicus clone CH230-166D21, WORKING DRAFT SEQUENCE.
 AC0989902.7 GI:30521241
 VERSION AC0989902.7
 KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
 SOURCE Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 221952)
 Muzny, D. Marie, Metzker, M. Lee, Abrazon, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsdorfs, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biewald, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Butrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M. L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Benson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Evans, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C. M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gargarescoris, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guayra, W.,
 Gunareeratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Haylak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hogues, M.,
 Hollins, B., Howells, S., Huix, J., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolyvet, A.,
 Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Kowis, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorenshuewa, L., Louiseged, H., Lozado, R. J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindaratne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawney, S., McLeod, M. P., McNeill, T. Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwakoilemeh, O., Okwomou, G., Olarnpunasagoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plöpper, F., Poindexter, A., Popovic, D., Prims, E., Pu, L.-L.,
 Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J.,
 Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
 Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajls, D.,
 Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
 Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C.,
 Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Uman, K.,
 Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J.,
 Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
 Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K.,
 Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
 Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhao, S., Dunn, D., von
 Niederhausern, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O.,
 Weinstein, G. and Gibbs, R. A.

TITLE
 JOURNAL
 Unpublished

REFERENCE 2 (bases 1 to 221952)
 AUTHORS Worley, K. C.
 TITLE Direct Submission
 JOURNAL Submitted (05-NOV-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 221952)
 Rattus norvegicus
 Rat Genome Sequencing Consortium.
 Submitted (10-MAY-2003) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On May 10, 2003 this sequence version replaced gi:25089200.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GINS
 Center clone name: CH230-166D21
 ----- Summary Statistics
 Assembly program: Atlas 3.0;
 Consensus quality: 213950 bases at least Q40
 Consensus quality: 215107 bases at least Q30
 Consensus quality: 215809 bases at least Q20
 Estimated insert size: 225832; sum-of-contigs estimation
 Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length
 (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 1 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 1 221952: contig of 221952 bp in length.
 Location/Qualifiers
 1. 221952
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-166D21"
 /complement(217689..218272)
 /note="clone_boundary
 clone end: 17
 site:ECORI
 end_sequence=BH279251"
 misc_feature
 misc_feature
 BASE COUNT 63742 a 47344 c 43902 g 61410 t 5554 others
 ORIGIN

Query Match 0.9%; Score 22; DB 2; Length 221952;
 Best Local Similarity 100.0%; Pred. No. 7.8;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

318 TTATCAAGTGAATGGAATA 339
 |||


```

DEFINITION Sequence 44 from Patent WO0198334.
ACCESSION AX343113
VERSION AX343113.1 GI:18152293
KEYWORDS
SOURCE unidentified
ORGANISM unidentified
REFERENCE 1
AUTHORS Hamel,D., Ouellet,C., Charland,N., Martin,D. and Brodeur,B.
TITLE Streptococcus antigens
JOURNAL Patent: WO 0198334-A 44 27-DEC-2001;
SHIRE BIOCHEM INC., (CA)
FEATURES
    source
        /organism="unidentified"
        /mol_type="genomic DNA"
        /db_xref="taxon:32644"
        /note="HAMU 264"
BASE COUNT 7 a 8 c 6 g 8 t
ORIGIN
Query Match 0.9%; Score 21; DB 6; Length 29;
Best Local Similarity 100.0%; Pred. No. 54;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 621 CTGCAGAACCTTCTCTATCTG 641
Db 9 CTGCAGAACCTTCTCTATCTG 29

RESULT 70
LOCUS SIWM7004AE 222 bp DNA linear VRL 03-AUG-1993
DEFINITION Simian immunodeficiency virus (M87004) glycoprotein 120 (env) gene.
ACCESSION M96201
VERSION M96201.1 GI:334607
KEYWORDS envelope-associated protein; glycoprotein 120.
SOURCE Simian immunodeficiency virus
ORGANISM Simian immunodeficiency virus
Viruses; Retrovird viruses; Retroviridae; Lentivirus; Primate
Lentivirus group.
1 (bases 1 to 222)
Overbaugh,J. and Rudensey,L.M.
Alterations in potential sites for glycosylation predominate during
evolution of the simian immunodeficiency virus envelope gene in
macaques
J. Virol. 66 (10), 5937-5948 (1992)
JOURNAL
MEDLINE 92407988
PUBMED 1527847
COMMENT Original source text: Simian immunodeficiency virus
(individual isolate Mne) DNA.
Location/Qualifiers
1..222
/organism="Simian immunodeficiency virus"
/proviral
/mol_type="genomic DNA"
/isolate="Mne"
/db_xref="taxon:11723"
BASE COUNT 45 a 40 c 40 g 97 t
ORIGIN
Query Match 0.9%; Score 21; DB 14; Length 222;
Best Local Similarity 100.0%; Pred. No. 46;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 757 CAACAACAGCACTACTACAG 777
Db 113 CAACAACAGCACTACTACAG 93

RESULT 71
LOCUS AF406703 2065 bp mRNA linear PLN 02-SEP-2002

```

```

DEFINITION Solanum tuberosum BE11-related homeotic protein 30 (Bel30) mRNA,
complete cds.
ACCESSION AF406703
VERSION AF406703.1 GI:22652126
KEYWORDS
SOURCE Solanum tuberosum (potato)
ORGANISM Solanum tuberosum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; Lamiales; Solanales; Solanaceae; Solanum.
REFERENCE 1
AUTHORS Chen,H., Rosin,F.M. and Hannapel,D.J.
TITLE A KNOX protein of potato interacts with several members of the TALE
family of transcription factors
Unpublished
2 (bases 1 to 2065)
Chen,H., Rosin,F.M. and Hannapel,D.J.
Direct Submission
Submitted (07-AUG-2001) Department of Horticulture, Iowa State
University, 257 Horticulture Hall, Ames, IA 50011-1100, USA
FEATURES
    source
        /organism="Solanum tuberosum"
        /mol_type="mRNA"
        /cultivar="Desiree"
        /db_xref="taxon:4113"
        /gene="Bel30"
        /gene="Bel30"
        /gene="Bel30"
        /note="StBel30"
        /codon_start=1
        /product="BE11-related homeotic protein 30"
        /protein_id="AA03627.1"
        /db_xref="GI:22652127"
        /translation="MATYFPSPNNQADQTFYFRGSLPESYSASNAPRMVFMN
YSSGAYSDMLTGTSSQOQHCIDIPISGATPNTSOQCLISNLGSGQWGIQDFSSWRD
SRNEMLANVNOVAVNOVGGLSLISGNIPTSGISHSVQSONPVQGGPMSFGSDG
NSQKQENADYFPDPNPRGLDMKGNVSGYSSIRRTIPSSYLYLAAOYLDEYV
SVRAIKKQNSKKEITKDSRSDVDKSISSDTPNAGSNPHESKNQSELSPTKOE
VQNKIAKLMSLDEIDRRYQYHQMQIVSSFVAVAGEGAKPYTALALQITSHFR
CLRDALICQIRASRSRLGEODASENSKAGISRLAFVPHDHLRQGLQOLGMQOHW
RPGRLPSSSVLRLAMLFHFHLPKDSKIMLAROTGLRSQVSMFIVARVLM
KPMVEWMEKASGAKKIDNSGSDVAPPLATKDSKVERGELHONAASEFEQVNGOI
LESXKNHADYEMEGASNAETROSQSGMNGEPLRPANDNCTLPDAVQGSNDRSEF
GSGSGNVLPNGVSLTGLQGBGSLPMSLTHSVYVPLRADMYSTAPTTWPETA
EFNCIDSGNRQOPFWLPSAT"
BASE COUNT 679 a 409 c 446 g 531 t
ORIGIN
Query Match 0.9%; Score 21; DB 8; Length 2065;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1102 TAAATAGCTCAATTCCTC 1122
Db 1433 TAAATAGCTCAATTCCTC 1453

RESULT 72
LOCUS AC014153 29183 bp DNA linear HTG 16-NOV-1999
DEFINITION Drosophila melanogaster. *** SEQUENCING IN PROGRESS ***
ACCESSION AC014153
VERSION AC014153.1 GI:6437182
KEYWORDS HTG; HTGS PHASE2.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1
AUTHORS Adams,M. and Venter,J.C.

```

TITLE Direct Submission
JOURNAL Submitted (16-NOV-1999) Celera Genomics, 45 West Gude Drive, Rockville, MD, USA

COMMENT This sequence was identified as CDM:10214207 by the submitter.
For further information on this sequence e-mail to fly@celera.com.
* NOTE: This is a 'working draft' sequence.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.

FEATURES
source
1..29183
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"

BASE COUNT 8559 a 5972 c 6050 g 8602 t

ORIGIN

Query Match 0.9%; Score 21; DB 2; Length 29183;
Best Local Similarity 100.0%; Pred. No. 32;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1508 AACCAATTCGAATTGAG 1528
|||||
Db 1881 AACCAATTCGAATTGAG 1861

RESULT 73
BX510940
LOCUS BX510940 58408 bp DNA linear HTG 17-MAY-2003
DEFINITION Dario rerio clone DKXP-69C1, *** SEQUENCING IN PROGRESS ***, 21
unordered pieces.
ACCESSION BX510940
VERSION BX510940.2 GI:30911000
KEYWORDS HTG; HTGS; PHASE1.
SOURCE Dario rerio (zebrafish)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 58408)
Sims, S.
REFERENCE Direct Submission
AUTHORS Submitted (16-MAY-2003) Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: zfish-help@sanger.ac.uk
JOURNAL zfish-help@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
COMMENT On May 19, 2003 this sequence version replaced gi:30840166.
----- Genome Center
Center: Wellcome Trust Sanger Institute
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: zfish-help@sanger.ac.uk
----- Project Information
Center project name: zkp69c1
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Chemistry: Dye-terminator; 100% of reads
Consensus quality: 34830 bases at least Q40
Consensus quality: 39888 bases at least Q30
Consensus quality: 46207 bases at least Q20
Insert size: 56408; sum-of-contents
Insert size: 186298; 2.8% error; agarose-fp
Quality coverage: 1.39x in Q20 bases; sum-of-contents Quality coverage: 0.90x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently
* consists of 21 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 2066: contig of 2066 bp in length

2067 2166: gap of 100 bp
* 2167 4260: contig of 2094 bp in length
* 4261 4360: gap of 100 bp
* 4361 6748: contig of 2388 bp in length
* 6749 6848: gap of 100 bp
* 6849 10662: contig of 3814 bp in length
* 10663 10762: gap of 100 bp
* 10763 13022: contig of 2260 bp in length
* 13023 13122: gap of 100 bp
* 13123 15364: contig of 2242 bp in length
* 15365 15464: gap of 100 bp
* 15465 20241: contig of 4777 bp in length
* 20242 20341: gap of 100 bp
* 20342 23411: contig of 3070 bp in length
* 23412 23511: gap of 100 bp
* 23512 26792: contig of 3281 bp in length
* 26793 26892: gap of 100 bp
* 26893 29362: contig of 2470 bp in length
* 29363 29462: gap of 100 bp
* 29463 32085: contig of 2623 bp in length
* 32086 32185: gap of 100 bp
* 32186 34199: contig of 2014 bp in length
* 34200 34299: gap of 100 bp
* 34300 36521: contig of 2222 bp in length
* 36522 36621: gap of 100 bp
* 36622 39482: contig of 2861 bp in length
* 39483 39582: gap of 100 bp
* 39583 41619: contig of 2037 bp in length
* 41620 41719: gap of 100 bp
* 41720 44437: contig of 2718 bp in length
* 44438 44537: gap of 100 bp
* 44538 47249: contig of 2712 bp in length
* 47250 47349: gap of 100 bp
* 47350 50165: contig of 2816 bp in length
* 50166 50265: gap of 100 bp
* 50266 53303: contig of 3038 bp in length
* 53304 53403: gap of 100 bp
* 53404 55413: contig of 2010 bp in length
* 55414 55513: gap of 100 bp
* 55514 58408: contig of 2895 bp in length.

FEATURES
source
1..58408
/organism="Dario rerio"
/mol_type="genomic DNA"
/db_xref="taxon:7955"
/clone="DKXP-69C1"
/clone_id="DarioKeyPilotC"
1..2066
/note="assembly fragment: 00679
fragment chain:1"
2167..4260
/note="assembly fragment: 00562
fragment chain:1"
4361..6748
/note="assembly fragment: 00982
fragment chain:2"
6849..10662
/note="assembly fragment: 00285
fragment chain:2"
10763..13022
/note="assembly fragment: 00075"
13123..15364
/note="assembly fragment: 00121"
15465..20241
/note="assembly fragment: 00128"
20342..23411
/note="assembly fragment: 00158"
23512..26792
/note="assembly fragment: 00208"
26893..29362
/note="assembly fragment: 00221"
29463..32085
/note="assembly fragment: 00292"

```

misc_feature      32186..34199
                   /note="assembly_fragment:00396"
misc_feature      34300..36521
                   /note="assembly_fragment:00426"
misc_feature      36622..39482
                   /note="assembly_fragment:00596"
misc_feature      39583..41619
                   /note="assembly_fragment:00867"
misc_feature      41720..44437
                   /note="assembly_fragment:00998"
misc_feature      44538..47249
                   /note="assembly_fragment:01004"
misc_feature      47350..50165
                   /note="assembly_fragment:01032"
misc_feature      50266..53303
                   /note="assembly_fragment:01088"
misc_feature      53404..55413
                   /note="assembly_fragment:01124"
misc_feature      55514..58408
                   /note="assembly_fragment:01319"
BASE COUNT      18075 a 10500 c 9854 g 17964 t 2015 others
ORIGIN
Query Match      0.9%; Score 21; DB 2; Length 58408;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      2368 TGTAGTAAAGAAAAATATA 2368
Db      54333 TGTAGTAAAGAAAAATATA 54353

RESULT 74
AC120849          70046 bp      DNA      linear      HTG 13-MAY-2002
DEFINITION      Mus musculus clone RP23-342010, LOW-PASS SEQUENCE SAMPLING.
ACCESSION      AC120849
VERSION      AC120849.1 GI:20531830
KEYWORDS      HTG; HTGS PHASE0.
SOURCE      Mus musculus (house mouse)
ORGANISM      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.
1 (bases 1 to 70046)
Mus musculus, clone RP23-342010
Unpublished
2 (bases 1 to 70046)
Birken, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
Bouhassira, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Charato, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatae, A., Kelle, C., Larocque, K., Lamaizares, R.,
Landers, T., Lenoczky, J., Levine, R., Lindblad-Toh, K., Liu, G.,
Maclean, C., MacDonald, P., Major, J., Margulis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Meneus, L.,
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Struss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
TITLE      Direct Submission
JOURNAL      Submitted (13-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

```

COMMENT

```

All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L19545
Center clone name: 342_O_10
-----
* NOTE: This record contains 86 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
1      713: contig of 713 bp in length
*      714      813: gap of 100 bp
*      814      1482: contig of 669 bp in length
*      1483      1583: gap of 100 bp
*      1583      2307: contig of 725 bp in length
*      2308      2407: gap of 100 bp
*      2408      3144: contig of 737 bp in length
*      3145      3244: gap of 100 bp
*      3245      3972: contig of 728 bp in length
*      3973      4072: gap of 100 bp
*      4073      4793: contig of 721 bp in length
*      4794      4893: gap of 100 bp
*      4894      5577: contig of 684 bp in length
*      5578      5677: gap of 100 bp
*      5678      6391: contig of 714 bp in length
*      6392      6491: gap of 100 bp
*      6492      7207: contig of 716 bp in length
*      7208      7307: gap of 100 bp
*      7308      8035: contig of 728 bp in length
*      8036      8135: gap of 100 bp
*      8136      8861: contig of 726 bp in length
*      8862      8961: gap of 100 bp
*      8962      9679: contig of 718 bp in length
*      9680      9779: gap of 100 bp
*      9780      10508: contig of 729 bp in length
*      10509      11321: gap of 100 bp
*      11322      11421: contig of 713 bp in length
*      11422      12126: contig of 705 bp in length
*      12127      12226: gap of 100 bp
*      12227      12928: contig of 702 bp in length
*      12929      13028: contig of 100 bp
*      13029      13735: gap of 707 bp in length
*      13736      13835: gap of 100 bp
*      13836      14556: contig of 721 bp in length
*      14557      14656: gap of 100 bp
*      14657      15382: contig of 726 bp in length
*      15383      15482: gap of 100 bp
*      15483      16309: contig of 727 bp in length
*      16310      16309: gap of 100 bp
*      16310      17045: contig of 736 bp in length
*      17046      17145: gap of 100 bp
*      17146      17874: contig of 729 bp in length
*      17875      17974: gap of 100 bp
*      17975      18704: contig of 730 bp in length
*      18705      18804: gap of 100 bp
*      18805      19522: contig of 718 bp in length
*      19523      19623: gap of 100 bp
*      19623      20307: contig of 685 bp in length
*      20307: gap of 100 bp
*      20407: gap of 100 bp

```

```

* 20408 21100: contig of 693 bp in length
* 21101 21200: gap of 100 bp
* 21201 21918: contig of 718 bp in length
* 21919 22018: gap of 100 bp
* 22019 22730: contig of 712 bp in length
* 22731 22830: gap of 100 bp
* 22831 23544: contig of 714 bp in length
* 23545 24378: contig of 734 bp in length
* 24379 24478: gap of 100 bp
* 24479 25209: contig of 731 bp in length
* 25210 25309: gap of 100 bp
* 25310 26041: contig of 732 bp in length
* 26042 26141: gap of 100 bp
* 26142 26870: contig of 723 bp in length
* 26871 26970: gap of 100 bp
* 26971 27699: contig of 725 bp in length
* 27700 27799: gap of 100 bp
* 27800 28508: contig of 709 bp in length
* 28509 28608: gap of 100 bp
* 28609 29316: contig of 708 bp in length
* 29317 29416: gap of 100 bp
* 29417 30139: contig of 723 bp in length
* 30140 30239: gap of 100 bp
* 30240 30961: contig of 722 bp in length
* 30962 31061: gap of 100 bp
* 31062 31751: contig of 699 bp in length
* 31752 31851: gap of 100 bp
* 31852 32568: contig of 717 bp in length
* 32569 32668: gap of 100 bp
* 32669 33382: contig of 714 bp in length
* 33383 34198: gap of 100 bp
* 34199 34298: contig of 716 bp in length
* 34299 35016: contig of 718 bp in length
* 35017 35116: gap of 100 bp
* 35117 35858: contig of 742 bp in length
* 35859 35958: gap of 100 bp
* 35959 36692: contig of 734 bp in length
* 36693 36792: gap of 100 bp
* 36793 37515: contig of 723 bp in length
* 37516 37615: gap of 100 bp
* 37616 38298: contig of 683 bp in length
* 38299 38398: gap of 100 bp
* 38399 39092: contig of 694 bp in length
* 39093 39192: gap of 100 bp
* 39193 39689: contig of 697 bp in length
* 39690 39989: gap of 100 bp
* 39990 40684: contig of 695 bp in length
* 40685 40784: gap of 100 bp
* 40785 41498: contig of 714 bp in length
* 41499 41598: gap of 100 bp
* 41599 42314: contig of 716 bp in length
* 42315 42414: gap of 100 bp
* 42415 43141: contig of 727 bp in length
* 43142 43241: gap of 100 bp
* 43242 43961: contig of 720 bp in length
* 43962 44061: gap of 100 bp
* 44062 44768: contig of 707 bp in length
* 44769 44868: gap of 100 bp
* 44869 45591: contig of 723 bp in length
* 45592 45691: gap of 100 bp
* 45692 46395: contig of 704 bp in length
* 46396 46495: gap of 100 bp
* 46496 47178: contig of 683 bp in length
* 47179 47278: gap of 100 bp
* 47279 48005: contig of 727 bp in length
* 48006 48105: gap of 100 bp
* 48106 48828: contig of 723 bp in length
* 48829 48928: gap of 100 bp
* 48929 49629: contig of 700 bp in length
* 49630 50446: gap of 100 bp
* 50446 50446: contig of 718 bp in length

```

```

* 50447 50546: gap of 100 bp
* 50547 51279: contig of 733 bp in length
* 51280 51379: gap of 100 bp
* 51380 52102: contig of 723 bp in length
* 52103 52202: gap of 100 bp
* 52203 52920: contig of 718 bp in length
* 52921 53020: gap of 100 bp
* 53021 53751: contig of 731 bp in length
* 53752 53851: gap of 100 bp
* 53852 54582: contig of 731 bp in length
* 54583 54682: gap of 100 bp
* 54683 55410: contig of 728 bp in length
* 55411 55510: gap of 100 bp
* 55511 56237: contig of 727 bp in length
* 56238 56337: gap of 100 bp

Query Match      0.9%; Score 21; DB 2; Length 70046;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy      757  |||||CAACACGACACTACAG 777
Db      69406 CACACACGACACTACAG 69426

RESULT 75
AC120849/c
LOCUS
DEFINITION
Mus musculus clone RP23-342010, LOW-PASS SEQUENCE SAMPLING.
AC120849
VERSION
AC120849.1 GI:20531830
KEYWORDS
HTG; HTGS PHASED.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 70046)
REFERENCE
Birren, B., Linton, L., Nusbaum, C. and Lander, E.
Mus musculus, clone RP23-342010
Unpublished
2 (bases 1 to 70046)
REFERENCE
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguski, L.,
Boukhalil, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Choquet, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., DeRellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Garde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Labrecque, K., Lamazares, R.,
Landers, T., Lepocky, J., Levine, R., Lindblad-Toh, K., Liu, G.,
Maclean, C., MacDonald, P., Major, J., Margus, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menus, L.,
Mhova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neil, D.,
Oliver, J., Peterson, K., Phunthang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Riback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupack, R.,
Seaman, S., Severy, P., Spencer, B., Strange-Thomann, N., Stojanovic, N.,
Strauss, K., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliou, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (13-MAY-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
All repeats were identified using RepeatMasker:
Snit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu

```

COMMENT

Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L19545
Center clone name: 342_O_10

* NOTE: This record contains 86 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1 713: contig of 713 bp in length
* 714 813: gap of 100 bp
* 814 1482: contig of 669 bp in length
* 1483 1582: gap of 100 bp
* 1583 2307: contig of 725 bp in length
* 2308 2407: gap of 100 bp
* 2408 3144: contig of 737 bp in length
* 3145 3244: gap of 100 bp
* 3245 3972: contig of 728 bp in length
* 3973 4072: gap of 100 bp
* 4073 4793: contig of 721 bp in length
* 4794 4893: gap of 100 bp
* 4894 5577: contig of 684 bp in length
* 5578 5677: gap of 100 bp
* 5678 6391: contig of 714 bp in length
* 6392 6491: gap of 100 bp
* 6492 7207: contig of 716 bp in length
* 7208 7307: gap of 100 bp
* 7308 8035: contig of 728 bp in length
* 8036 8135: gap of 100 bp
* 8136 8861: contig of 726 bp in length
* 8862 8961: gap of 100 bp
* 8962 9679: contig of 718 bp in length
* 9680 9779: gap of 100 bp
* 9780 10508: contig of 729 bp in length
* 10509 10608: gap of 100 bp
* 10609 11321: contig of 713 bp in length
* 11322 11421: gap of 100 bp
* 11422 12126: contig of 705 bp in length
* 12127 12226: gap of 100 bp
* 12227 12928: contig of 702 bp in length
* 12929 13028: gap of 100 bp
* 13029 13735: contig of 707 bp in length
* 13736 13835: gap of 100 bp
* 13836 14556: contig of 721 bp in length
* 14557 14656: gap of 100 bp
* 14657 15382: contig of 726 bp in length
* 15383 15482: gap of 100 bp
* 15483 16209: contig of 727 bp in length
* 16210 16309: gap of 100 bp
* 16310 17045: contig of 736 bp in length
* 17046 17145: gap of 100 bp
* 17146 17874: contig of 729 bp in length
* 17875 17974: gap of 100 bp
* 17975 18704: contig of 730 bp in length
* 18705 18804: gap of 100 bp
* 18805 19522: contig of 718 bp in length
* 19523 19622: gap of 100 bp
* 19623 20307: contig of 685 bp in length
* 20308 20407: gap of 100 bp
* 20408 21100: contig of 693 bp in length
* 21101 21200: gap of 100 bp
* 21201 21918: contig of 718 bp in length
* 21919 22018: gap of 100 bp
* 22019 22730: contig of 712 bp in length
* 22731 22830: gap of 100 bp
* 22831 23544: contig of 714 bp in length

* 23545 23644: gap of 100 bp
* 23645 24378: contig of 734 bp in length
* 24379 24478: gap of 100 bp
* 24479 25209: contig of 731 bp in length
* 25210 25309: gap of 100 bp
* 25310 26041: contig of 732 bp in length
* 26042 26141: gap of 100 bp
* 26142 26870: contig of 729 bp in length
* 26871 26970: gap of 100 bp
* 26971 27599: contig of 729 bp in length
* 27600 27799: gap of 100 bp
* 27800 28508: contig of 709 bp in length
* 28509 29116: contig of 708 bp in length
* 29117 29416: gap of 100 bp
* 29417 30139: contig of 723 bp in length
* 30140 30239: gap of 100 bp
* 30240 30961: contig of 722 bp in length
* 30962 31061: gap of 100 bp
* 31062 31751: contig of 690 bp in length
* 31752 31851: gap of 100 bp
* 31852 32568: contig of 717 bp in length
* 32569 32668: gap of 100 bp
* 32669 33382: contig of 714 bp in length
* 33383 33482: gap of 100 bp
* 33483 34198: contig of 716 bp in length
* 34199 34299: gap of 100 bp
* 34299 35016: contig of 718 bp in length
* 35017 35116: gap of 100 bp
* 35117 35858: contig of 742 bp in length
* 35859 35958: gap of 100 bp
* 35959 36692: contig of 734 bp in length
* 36693 36792: gap of 100 bp
* 36793 37515: contig of 723 bp in length
* 37516 37615: gap of 100 bp
* 37616 38298: contig of 683 bp in length
* 38299 38398: gap of 100 bp
* 38399 39092: contig of 694 bp in length
* 39093 39192: gap of 100 bp
* 39193 39889: contig of 697 bp in length
* 39890 39989: gap of 100 bp
* 39990 40684: contig of 695 bp in length
* 40685 40784: gap of 100 bp
* 40785 41498: contig of 714 bp in length
* 41499 41598: gap of 100 bp
* 41599 42314: contig of 716 bp in length
* 42315 42414: gap of 100 bp
* 42415 43141: contig of 727 bp in length
* 43142 43241: gap of 100 bp
* 43242 43961: contig of 720 bp in length
* 43962 44061: gap of 100 bp
* 44062 44768: contig of 707 bp in length
* 44769 44868: gap of 100 bp
* 44869 45591: contig of 723 bp in length
* 45592 45691: gap of 100 bp
* 45692 46395: contig of 704 bp in length
* 46396 46495: gap of 100 bp
* 46496 47178: contig of 683 bp in length
* 47179 47278: gap of 100 bp
* 47279 48005: contig of 727 bp in length
* 48006 48105: gap of 100 bp
* 48106 48828: contig of 723 bp in length
* 48829 48928: gap of 100 bp
* 48929 49628: contig of 700 bp in length
* 49629 49728: gap of 100 bp
* 49729 50446: contig of 718 bp in length
* 50447 50546: gap of 100 bp
* 50547 51279: contig of 733 bp in length
* 51280 51379: gap of 100 bp
* 51380 52102: contig of 723 bp in length
* 52103 52202: gap of 100 bp
* 52203 52920: contig of 718 bp in length
* 52921 53020: gap of 100 bp

```

*      53021      53751: contig of 731 bp in length
*      53752      53851: gap of 100 bp
*      53852      54582: contig of 731 bp in length
*      54583      54682: gap of 100 bp
*      54683      55410: contig of 728 bp in length
*      55411      55510: gap of 100 bp
*      55511      56237: contig of 727 bp in length
*      56238      56337: gap of 100 bp

Query Match      0.9%; Score 21; DB 2; Length 70046;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      757      CAACACAGCACTACTAAG 777
Db      66124      CAACACAGCACTACTAAG 66104

RESULT 76
LOCUS      AL731563      129559 bp      DNA      linear      PRI 03-JUN-2002
DEFINITION      Human DNA sequence from clone Rpl1-344N10 on chromosome 10,
complete sequence.
ACCESSION      AL731563 AC024948
VERSION        AL731563.9 GI:21322403
KEYWORDS      HTG.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 129559)
Whitehead,S.
REFERENCE      Submitted (01-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
AUTHORS      Cambridgehire, CB10 1SA, UK. E-mail enquiries:
TITLE      humquery@sanger.ac.uk Clone requests: clonequest@sanger.ac.uk
JOURNAL      On Jun 4, 2002 this sequence version replaced gi:20520536.
COMMENT      Draft Sequence Produced by Genome Therapeutics Corp, 100 Beaver
Street, Waltham, MA 02453, USA
http://www.genomecorp.com
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Emi, EMBL; Swi,
SWISSPROT; Tr, TrEMBL; Wp, WormPEP; Information on the WormPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
Rpl1-344N10 is from the library RPCI-11.2 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pACE3.6.
location/Qualifiers
1..129559
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="Rpl1-344N10"

```

```

BASE COUNT      41513 a 25955 c 24645 g 37446 t
ORIGIN
Query Match      0.9%; Score 21; DB 9; Length 129559;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1599      ATGAACATGATATATCAGTG 1619
Db      113842      ATGAACATGATATATCAGTG 113862

RESULT 77
LOCUS      AC005550      133237 bp      DNA      linear      PRI 03-FEB-2000
DEFINITION      Homo sapiens PAC clone RPA-620P6 from 7p21-p22, complete sequence.
ACCESSION      AC005550
VERSION        AC005550.1 GI:3478668
KEYWORDS      HTG.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 133237)
Andrews,S., Lacy,M. and Dubbelde,C.
REFERENCE      The sequence of Homo sapiens PAC clone RPA-620P6
AUTHORS      Unpublished
TITLE      2 (bases 1 to 133237)
JOURNAL      Waterston,R.
REFERENCE      Direct Submission
AUTHORS      Submitted (27-AUG-1998) Department of Genetics, Washington
TITLE      University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
JOURNAL      3 (bases 1 to 133237)
COMMENT      ----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wuston.wustl.edu
----- Summary Statistics
Center project name: H_DJ0620P06

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see
http://www.nhgri.nih.gov/DIR/GRB/CHR7, send
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-4, prepared by Pieter de Jong and coworkers at Roswell Park Cancer Institute, using the method described by Ioannou et al., Nature Genetics

6:84-9 (1994). The library is from one male donor. For further details, see <http://bacpac.med.buffalo.edu/>. The clone is available from Genome Systems, Inc. (<http://www.genomesystems.com>).
 VECTOR: PCYPAC2
 NEIGHBORING SEQUENCE INFORMATION:
 The clone sequenced to the left is RP4-799L20, 200 bp overlap. Actual start of this clone is at base position 197 of RP4-620P6; actual end is at 133237 of RP4-620P6.

FEATURES

```

SOURCE          1..133237
                  Location/Qualifiers
                    1..133237
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="7"
                    /map="7p21-p22"
                    /clone="RP4-620P6"
                    /clone_11b="RP4-620P6"
                    /gene="WUGSC:H_DJ0620P06.1"
                    /complement(joinr<921..60577)
                    /gene="WUGSC:H_DJ0620P06.1"
                    /complement(joinr<921..1093,60061..60577)
                    /note="homeobox protein mox-2; match to P50222
                    (PID:g1709079); H_DJ0620P06.1"
                    /codon_start=1
                    /evidence=not experimental
                    /protein_id="AAC33152.1"
                    /db_xref="GI:3478669"
                    /translation="MEHPFGCLRSPHATQGLHPSSQSLALHGRSDHMSYPELST
                    SSSCIAGYPNEEGMFASQHHGHHHHHHHHHHHHHHHHHQQOQALQTNMFLPQMSPPS
                    AARHSLCLDPDGGPELSPPLVCSNSSISGSPPTPAACAPGDPYGQALSPAEAE
                    KRSGGKRSDDSDSOEGNYKSEVNSKPKRKERTAFKQELRELEAFHNNYTRLRVY
                    EAVVNLDLTERQ"
misc_feature     2336..2759
                  /note="match to EST AA449360 (NID:g2163209) zx08e03.sl"
                  /complement(2910..3387)
                  /gene="WUGSC:H_DJ0620P06.1"
                  /note="match to EST AA449620 (NID:g2163370) zx08e03.r1"
repeat_region    8366..8409
                  /rpt_family="MER1_type"
repeat_region    8468..8554
                  /rpt_family="MER1_type"
repeat_region    8560..8738
                  /rpt_family="MER53"
repeat_region    10864..11209
                  /rpt_family="MALR"
repeat_region    12676..15454
                  /rpt_family="L1"
repeat_region    15453..17195
                  /rpt_family="L1"
repeat_region    17182..17283
                  /rpt_family="L1"
repeat_region    17309..17603
                  /rpt_family="L1"
repeat_region    24272..24472
                  /rpt_family="L2"
repeat_region    24741..24870
                  /rpt_family="L2"
repeat_region    25671..25868
                  /rpt_family="L2"
repeat_region    26843..32728
                  /rpt_family="L1"
repeat_region    32747..33102
                  /rpt_family="MALR"
repeat_region    33103..33169
                  /rpt_family="L1"
repeat_region    33182..33373
                  /rpt_family="Alu"
repeat_region    33654..33953
                  /rpt_family="Alu"
repeat_region    34252..34387
                  /rpt_family="L2"
repeat_region    34607..34825

```

```

repeat_region    /rpt_family="MER1_type"
                  34916..35207
repeat_region    /rpt_family="Alu"
                  37509..37813
repeat_region    /rpt_family="Alu"
                  39217..39260
repeat_region    /rpt_family="(CA)n"
                  40647..40946
repeat_region    /rpt_family="Alu"
                  40797..40947
misc_feature     /note="similar to EST T40302 (NID:g647938) ya31g04.r5"
                  41185..41484
repeat_region    /rpt_family="Alu"
                  42287..42648
repeat_region    /rpt_family="MER2_type"
                  42656..42950
repeat_region    /rpt_family="Alu"
                  42951..43047
repeat_region    /rpt_family="MER2_type"
                  44678..44864
repeat_region    /rpt_family="MER1_type"
                  45060..45125
repeat_region    /rpt_family="MER1_type"
                  45950..46287
repeat_region    /rpt_family="Alu"
                  46786..47086
repeat_region    /rpt_family="Alu"
                  47616..47740
repeat_region    /rpt_family="L1"
                  50764..51047
repeat_region    /rpt_family="Alu"
                  51437..51496
repeat_region    /rpt_family="MIR"
                  55485..55884
misc_feature     /note="match to EST H82632 (NID:g1060721) ys70e10.r1"
                  56100..56440
misc_feature     /note="match to EST H95711 (NID:g1108853) yc95h12.s1"
                  /complement(56457..56682)
                  /gene="WUGSC:H_DJ0620P06.1"
misc_feature     /note="match to EST H95710 (NID:g1108852) yc95h12.r1"
                  58819..58913
repeat_region    /rpt_family="MIR"
                  /complement(60061..60844)
                  /db_xref="GI:1254905"
misc_feature     /complement(60324..60653)
                  /note="match to EST AA436934 (NID:g2141848) zv72b07.r1"
misc_feature     /complement(60329..60653)
                  /note="match to EST AA436934 (NID:g2141848) zv72b07.r1"
misc_feature     /complement(62208..62539)
                  /note="match to EST AA973922 (NID:g3149102) oo48g09.sl"
misc_feature     /complement(63250..63404)
                  /note="match to EST F06363 (NID:g671934)"
misc_feature     /complement(64871..65031)
                  /note="match to EST F06363 (NID:g671934)"
repeat_region    68732..68978
                  /rpt_family="MER1_type"
misc_feature     69255..69595
                  /note="match to EST AA478691 (NID:g2207325) zv09b01.r1"
repeat_region    70227..70409
                  /rpt_family="MIR"
repeat_region    70675..71141
                  /rpt_family="MER51B"
repeat_region    72265..72557
                  /rpt_family="Alu"
STS              74705..74929
                  /db_xref="GI:485254"
repeat_region    75009..75300
                  /rpt_family="Alu"
misc_feature     75560..75629

```

Query Match 0.9%; Score 21; DB 9; Length 133237;
 Best Local Similarity 100.0%; Pred.No. 28;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy      2319 CAGAAGCAGAAAAATTACTTG 2319
|       |||
Db      96040 CAGAAGCAGAAAAATTACTTG 96060

RESULT 78
LOCUS      AC010907      139357 bp      DNA      linear      PRI 09-JAN-2002
DEFINITION Homo sapiens BAC clone RP11-568H24 from 2, complete sequence.
ACCESSION  AC010907
VERSION     AC010907.10  GI:15321567
KEYWORDS
SOURCE
ORGANISM   Homo sapiens (human)
            Homo sapiens
            Eumalaya: Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE  1 (bases 1 to 139357)
            Sulston, J.E. and Waterston, R.
            Toward a complete human genome sequence
            Genome Res. 8 (11), 1097-1108 (1998)
JOURNAL     99063792
MEDLINE     9847074
PUBMED
REFERENCE  2 (bases 1 to 139357)
            Du, H., Haekenson, W. and Dixon, R.
            The sequence of Homo sapiens BAC clone RP11-568H24
            Unpublished (2001)
JOURNAL
REFERENCE  3 (bases 1 to 139357)
            Waterston, R.H.
            Direct Submission
            Submitted (25-SEP-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
JOURNAL     4 (bases 1 to 139357)
            Waterston, R.H.
            Direct Submission
            Submitted (28-AUG-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
JOURNAL     5 (bases 1 to 139357)
            Waterston, R.
            Direct Submission
            Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108,
            On Aug 28, 2001 this sequence version replaced gi:13399437.
COMMENT    ----- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: MUGSC
            Web site: http://genome.wustl.edu/gsc
            Contact: saplens@watsn.wustl.edu
            ----- Summary Statistics
            -----
            Center project name: H_NH0568H24

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McHerson, Department of Genetics, Washington University, St. Louis,
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

```

SOURCE INFORMATION: The Rpci-11 human BAC library was made from the blood of one male donor, as described by Oosagawa, K., Moon, P. Y., Zhao, B., Fringsen, E., Tateno, M., Caranese, J.J. and de Jong, P.J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACe3.6

NEIGHBORING SEQUENCE INFORMATION: The clone sequenced to the right is RP11-1785E, 2000 bp overlap. The clone sequenced to the left is at base position 1 of RP11-568H24. Actual start of this clone is at base position 1 of RP11-568H24.

The sequence between 66033 to 66578 and 104506 to 104590 is covered only by PCR products from clone DNA. The sequence contains a dinucleotide (TG) run from 65513 to 65634 in which the exact length is unknown. The sequence contains a dinucleotide (TC) run from 104366 bp to 104631 bp in which the exact length is unknown. The sequence from base position 4458 to 6187 can not be guaranteed due to a tandem repeat.

FEATURES

```
source
1..13937
/oranism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="caxon:9606"
/chromosome="2"
/map="2"
/clone="RP11-568H24"
/clone_lib="RPC1-11"
383..409
/rpt_family="AT_rich"
876..950
/rpt_family="ACHobo"
/rpt_family="Alu"
951..1248
/rpt_family="Alu"
1249..1339
/rpt_family="ACHobo"
1340..1514
/rpt_family="MERL_type"
1511..1965
/rpt_family="ACHobo"
1973..2279
/rpt_family="Alu"
2324..2843
/rpt_family="LI"
2844..3050
/rpt_family="MERL_type"
2943..2958
/note="similar to Homo sapiens EST A1597790 (NID:g4606838)
tr92n02.x1"
3051..3174
/rpt_family="LI"
3175..3483
/rpt_family="Alu"
3182..3206
/note="similar to Homo sapiens EST BG178637
(NID:g12685340)"
3484..4227
/rpt_family="LI"
4255..4330
/rpt_family="(T)n"
4312..4987
/rpt_family="SVA"
4571..6524
/note="CpG island (%GC=74.6, o/e=0.80, #CGs=159)"
4978..5152
/rpt_family="(CCGG)n"
5163..5560
/rpt_family="SVA"
5249..5273
/rpt_family="GC-rich"
5593..6486
```

```

repeat_region /rpt_family="SVA"
6443..6529
/rpt_family="GA-rich"
repeat_region 6531..6732
/rpt_family="L1"
repeat_region 6733..7033
/rpt_family="Alu"
repeat_region 7034..7249
/rpt_family="L1"
repeat_region 7250..7554
/rpt_family="Alu"
repeat_region 7555..8490
/rpt_family="L1"
repeat_region 8528..8726
/rpt_family="L1"
repeat_region 8666..8669
/rpt_family="ACHOB"
misc_feature /note="match to EST BF314952 (NID:g11263123)"
9031..9360
/note="match to EST BF314952 (NID:g11263123)"
misc_feature 9047..9360
/note="match to EST BF259985 (NID:g130920)"
9068..9069
/note="match to EST BF316717 (NID:g11265074)"
9068..9069
/note="match to EST BF198782 (NID:g14653803)"
9144..9147
/note="match to EST BF311850 (NID:g11259616)"
9239..9360
/note="match to EST BF198831 (NID:g14653852)"
9239..9360
/note="match to EST BF198308 (NID:g14653329)"
9239..9360
/note="match to EST BF198502 (NID:g14653523)"
9240..9360
/note="match to EST BF206235 (NID:g1109821)"
9240..9360
/note="match to EST BF311988 (NID:g11259763)"
9240..9360
/note="match to EST BF199896 (NID:g14654917)"
9240..9360
/note="match to EST BF313026 (NID:g11260952)"
9313..9360
/note="match to EST BF311981 (NID:g11259754)"
9313..9360
/note="match to EST BF316085 (NID:g11264380)"
9313..9360
/note="match to EST BF198253 (NID:g1465314)"
9322..9360
/note="match to EST BF316717 (NID:g11265074)"
9322..9360
/note="match to EST BF198782 (NID:g14653803)"
9322..9360
/note="match to EST BF311850 (NID:g11259616)"
9333..9360
/note="match to EST BF313758 (NID:g11264050)"

```

Query Match 0.9%; Score 21; DB 9; Length 139557;
 Best Local Similarity 100.0%; Pred. NO. 28;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1755 CAGATGTTAAAGCAATCCAA 1775
 |||||
 47594 CAGATGTTAAAGCAATCCAA 47614

RESULT 79

```

AC102254 146690 bp DNA linear HTG 21-AUG-2002
LOCUS Mus musculus clone RP24-181A10, WORKING DRAFT SEQUENCE, 15
DEFINITION unordered pieces.
ACCESSION AC102254.2 GI:22381162
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 146690)
Birken, B., Nusbaum, C. and Lander, E.
Mus musculus, clone RP24-181A10
Unpublished
2 (bases 1 to 146690)
Birken, B., Linton, L., Nusbaum, C., Lander, E., All, A., Allen, N.,
Anderson, S., Barina, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,
Brown, A., Camarata, A., Campopiano, A., Chang, J., Chazaro, B.,
Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,
Cooke, P., Deyrellano, K., Dewar, K., Diaz, J.S., Dodge, S., Faro, S.,
Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Heatford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,
Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K.,
Lamacz, R., Landers, T., Lehotzky, J., Levine, R., Liu, G.,
Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
McCarthy, M., McEwan, P., McKernan, K., McNeelers, R., Meldrum, J.,
Meneus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
Nobdu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
Oliwer, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
Roman, J., Rosetti, P., Roy, A., Santos, R., Schauer, S., Schupack, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Triggilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 146690)
Birken, B., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B.,
Barina, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepe, Y., Collymore, A.,
Cook, A., Cooke, P., Deyrellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Nobdu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schupack, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Testaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (21-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced g1117061340.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information

```

```

Center project name: L18267
Center clone name: 181.A.10
----- Summary Statistics -----
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 142759 bases at least Q40
Consensus quality: 144340 bases at least Q30
Consensus quality: 144937 bases at least Q20
Insert size: 151000; agarose-fp
Insert size: 145290; sum-of-contigs
Quality coverage: 7.1 in Q20 bases; agarose-fp
Quality coverage: 7.4 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 15 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1191: contig of 1191 bp in length
*
1192 1291: gap of 100 bp
*
1292 2633: contig of 1342 bp in length
*
2634 2733: gap of 100 bp
*
2734 5384: contig of 2651 bp in length
*
5385 5484: gap of 100 bp
*
5485 5698: contig of 4214 bp in length
*
5699 9798: gap of 100 bp
*
9799 14648: contig of 4850 bp in length
*
14649 14748: gap of 100 bp
*
14749 19191: contig of 4443 bp in length
*
19192 19291: gap of 100 bp
*
19292 24825: contig of 5534 bp in length
*
24826 24925: gap of 100 bp
*
24926 30955: contig of 6030 bp in length
*
30956 31055: gap of 100 bp
*
31056 37621: contig of 6566 bp in length
*
37622 37721: gap of 100 bp
*
37722 47762: contig of 10041 bp in length
*
47763 47863: gap of 100 bp
*
47864 62867: contig of 15005 bp in length
*
62868 62967: gap of 100 bp
*
62968 78984: contig of 16017 bp in length
*
78985 79084: gap of 100 bp
*
79085 100397: contig of 21313 bp in length
*
100398 100497: gap of 100 bp
*
100498 123488: contig of 22991 bp in length
*
123489 123589: gap of 100 bp
*
123590 146690: contig of 23102 bp in length.
*
FEATURES
Source
1. 146690
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone="RP24-181A10"
/clone.lib="RP24-181A10"
Location/Qualifiers
1. 1191
/feature="assembly_fragment"
1292..2633
/feature="assembly_fragment"
2734..5384
/feature="assembly_fragment"
5485..5698
/feature="assembly_fragment"
5799..14648
/feature="assembly_fragment"
14749..19191
/feature="assembly_fragment"
19292..24825
/feature="assembly_fragment"
24926..30955
/feature="assembly_fragment"
misc_feature
/feature="assembly_fragment"
31056..37621
/feature="assembly_fragment"
37722..47762
/feature="assembly_fragment"
47863..62867
/feature="assembly_fragment"
62968..78984
/feature="assembly_fragment"
79085..100397
/feature="assembly_fragment"
100498..123488
/feature="assembly_fragment"
123589..146690
/feature="assembly_fragment"
misc_feature
/feature="assembly_fragment"
45025 a 28279 c 28361 g 43625 c 1400 others
BASE COUNT
ORIGIN
Query Match 0.9%; Score 21; DB 2; Length 146690;
Best Local Similarity 100.0%; Pred.No. 28;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 264 CAAACTATAGCTAAAGATG 284
Db 111844 CAAACTATAGCTAAAGATG 111864
RESULT 80
AC104074 148348 bp DNA linear HTG 03-DEC-2001
LOCUS Homo sapiens chromosome UNK clone RP11-37004, WORKING DRAFT
DEFINITION
ACCESSION AC104074.1 GI:17227276
VERSION HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 148348)
Waterston,R.H.
The sequence of Homo sapiens clone
unpublished
2 (bases 1 to 148348)
Waterston,R.H.
Direct Submission
Submitted (03-DEC-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUSGC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: H_NH0370G04
----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: Plasmid; 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 126050 bases at least Q40
Consensus quality: 130849 bases at least Q30
Consensus quality: 134078 bases at least Q20
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 45 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.

```

* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1575: contig of 1575 bp in length
* 1576: gap of unknown length
* 1576: contig of 1187 bp in length
* 2963: gap of unknown length
* 2963: contig of 1055 bp in length
* 4018: gap of unknown length
* 4118: contig of 2162 bp in length
* 6280: gap of unknown length
* 6380: contig of 1709 bp in length
* 8089: gap of unknown length
* 8189: contig of 1738 bp in length
* 9297: gap of unknown length
* 10026: contig of 1274 bp in length
* 11007: gap of unknown length
* 11301: contig of 1236 bp in length
* 11401: gap of unknown length
* 12637: contig of 2128 bp in length
* 12737: gap of unknown length
* 14864: contig of 2128 bp in length
* 14865: gap of unknown length
* 14965: contig of 2171 bp in length
* 17136: gap of unknown length
* 17236: contig of 1471 bp in length
* 17236: gap of unknown length
* 18707: contig of 1727 bp in length
* 18807: gap of unknown length
* 20533: contig of 1515 bp in length
* 20534: gap of unknown length
* 22148: contig of 1709 bp in length
* 22249: gap of unknown length
* 22249: contig of 1709 bp in length
* 23957: gap of unknown length
* 24057: contig of 1910 bp in length
* 24058: gap of unknown length
* 25967: contig of 1746 bp in length
* 26067: gap of unknown length
* 26068: contig of 1746 bp in length
* 27814: gap of unknown length
* 27914: contig of 1638 bp in length
* 29551: gap of unknown length
* 29552: contig of 2598 bp in length
* 32249: gap of unknown length
* 32250: contig of 2331 bp in length
* 32350: gap of unknown length
* 34681: gap of unknown length
* 34780: contig of 2134 bp in length
* 36914: gap of unknown length
* 37015: contig of 1798 bp in length
* 38813: gap of unknown length
* 38913: contig of 1943 bp in length
* 38913: gap of unknown length
* 40856: gap of unknown length
* 40956: contig of 1824 bp in length
* 42780: gap of unknown length
* 42880: contig of 2226 bp in length
* 45106: gap of unknown length
* 45206: contig of 2185 bp in length
* 47391: gap of unknown length
* 47490: contig of 2720 bp in length
* 50221: gap of unknown length
* 50320: contig of 2827 bp in length
* 50321: gap of unknown length
* 53147: contig of 2114 bp in length
* 53248: gap of unknown length
* 53461: gap of unknown length
* 55462: contig of 2642 bp in length
* 58104: gap of unknown length
* 58203: contig of 4103 bp in length
* 62306: gap of unknown length
* 62406: contig of 3909 bp in length
* 62407: gap of unknown length
* 66316: gap of unknown length
* 66416: contig of 4260 bp in length
* 70675: gap of unknown length
* 70776: contig of 3831 bp in length
* 74606: gap of unknown length
* 74706: contig of 5199 bp in length
* 79905: gap of unknown length
* 80005: contig of 4404 bp in length
* 84409: gap of unknown length
* 84509: gap of unknown length

FEATURES
source
1. 148348
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="UNK"
/clone="RP11-370C4"
1. 1575
/note="assembly_name:Contig19"
1676. 2862
/note="assembly_name:Contig21"
2963. 4017
/note="assembly_name:Contig22"
4118. 6279
/note="assembly_name:Contig23"
6380. 8088
/note="assembly_name:Contig24"
8189. 9926
/note="assembly_name:Contig25"
10027. 11300
/note="assembly_name:Contig28"
11401. 12636
/note="assembly_name:Contig29"
12737. 14864
/note="assembly_name:Contig30"
14965. 17135
/note="assembly_name:Contig31"
17236. 18706
/note="assembly_name:Contig32"
18807. 20533
/note="assembly_name:Contig33"
20634. 22148
/note="assembly_name:Contig34"
22249. 23957
/note="assembly_name:Contig35"
24058. 25967
/note="assembly_name:Contig36"
26068. 27813
/note="assembly_name:Contig37"
27914. 29551
/note="assembly_name:Contig38"
29652. 32249
/note="assembly_name:Contig39"
32350. 34680
/note="assembly_name:Contig40"
34781. 36914
/note="assembly_name:Contig41"
37015. 38812
/note="assembly_name:Contig42"
38913. 40855
/note="assembly_name:Contig43"
40956. 42779
/note="assembly_name:Contig44"
42880. 45105
misc_feature
84510 88836: contig of 4327 bp in length
* 88937 88936: gap of unknown length
* 88937 95013: contig of 6077 bp in length
* 95014 95113: gap of unknown length
* 95114 99706: contig of 4593 bp in length
* 99707 99806: gap of unknown length
* 99807 104423: contig of 4617 bp in length
* 104424 104523: gap of unknown length
* 104524 109828: contig of 5305 bp in length
* 109829 109928: gap of unknown length
* 109929 114486: contig of 4558 bp in length
* 114487 114586: gap of unknown length
* 114587 120622: contig of 6036 bp in length
* 120623 120722: gap of unknown length
* 120723 127937: contig of 7215 bp in length
* 127938 128037: gap of unknown length
* 128038 137966: contig of 9929 bp in length
* 137967 138066: gap of unknown length
* 138067 148348: contig of 10282 bp in length.
Location/Qualifiers
1. 148348
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="UNK"
/clone="RP11-370C4"

```

misc_feature      /note="assembly_name:Contig45"
                  45206..47390
                  /note="assembly_name:Contig46"
misc_feature      47491..50220
                  /note="assembly_name:Contig47"
misc_feature      50321..53147
                  /note="assembly_name:Contig48"
misc_feature      53248..55361
                  /note="assembly_name:Contig49"
misc_feature      55462..58103

Query Match      0.9%; Score 21; DB 2; Length 146348;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1606 TGATTTATTCAGTGTGTAAGG 1626
Db      111749 TGATATATTCAGTGTGTAAGG 111769

RESULT 81
AL512380/c      AL512380      150972 bp      DNA      linear      PRI 13-MAR-2002
LOCUS      Human DNA sequence from clone RP11-204E9 on chromosome 6, complete
DEFINITION
ACCESSION      AL512380
VERSION
KEYWORDS
SOURCE      HTG.
ORGANISM      Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS      Corby,N.
TITLE      Direct Submission
JOURNAL
COMMENT      Submitted (12-MAR-2002) Wellcome Trust Sanger Institute, Hinxton,
                  Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
                  humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
                  On Mar 21, 2002 this sequence version replaced gi:16973010.
                  During sequence assembly data is compared from overlapping clones.
                  Where differences are found these are annotated as variations
                  together with a note of the overlapping clone name. Note that the
                  variation annotation may not be found in the sequence submission
                  corresponding to the overlapping clone, as we submit sequences with
                  only a small overlap as described above.
                  This sequence was finished as follows unless otherwise noted: all
                  regions were either double-stranded or sequenced with an alternate
                  chemistry or covered by high quality data (i.e., phred quality >=
                  30); an attempt was made to resolve all sequencing problems, such
                  as compressions and repeats; all regions were covered by at least
                  one plasmid subclone or more than one M13 subclone; and the
                  assembly was confirmed by restriction digest. The following
                  abbreviations are used to associate primary accession numbers given
                  in the feature table with their source databases: Em., EMBL; Sw.,
                  SWISSPROT; Tr., TrEMBL; Wp., WormPEP; Information on the WormPEP
                  database can be found at
                  http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
                  was generated from part of bacterial clone contigs of human
                  chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
                  Group. Further information can be found at
                  http://www.sanger.ac.uk/HGP/Chr6
                  RP11-204E9 is from the library RPCI-11.1 constructed by the group
                  of Pieter de Jong. For further details see
                  http://www.chori.org/bacpac/home.htm
                  VECTOR: DBAC63.6
IMPORTANT: This sequence is not the entire insert of clone
RP11-204E9 it may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true left end of clone RP1-32214 is at 148973 in this sequence.
The true right end of clone RP1-135122 is at 2000 in this sequence.
Location/Qualifiers
1..150972
/organism="Homo sapiens"

```

```

/mol type="genomic DNA"
/db xref="taxon:9606"
/chromosome="6"
/clone="RP11-204E9"
/clone_id="RPCI-11.1"
171..261
/note="Sequence from overlapping clone RP11-38C16
(AL035604). Assembly confirmed by restriction digest
data."
13315..13482
/note="Single clone region. Sequence from reads from a
short insert library derived from a clone PCR. Restriction
digest data confirm the assembly."
13447
/note="Tandem repeat. Forced join. Assembly confirmed by
restriction digest data."
21135
/note="Tandem repeat. Forced join. Gap size estimated to
be approximately 500bp by restriction digest data."
123937..124025
/note="Single clone region. Sequence from reads from a
short insert library derived from a single pUC clone.
Restriction digest data confirm the assembly."
complement(124552..124593)
/note="Single clone region. Sequence from reads from a
short insert library derived from a single pUC clone.
Restriction digest data confirm the assembly."

BASE COUNT      42379 a 31996 c 31132 g 45465 t
ORIGIN

Query Match      0.9%; Score 21; DB 9; Length 150972;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      54 TTTCCTATATGATGAAAC 74
Db      39019 TTTCCTATATGATGAAAC 38999

RESULT 82
AC020977/c      AC020977      156105 bp      DNA      linear      PRI 06-SEP-2001
LOCUS      Homo sapiens chromosome 5 clone RP11-92M7, complete sequence.
DEFINITION
ACCESSION      AC020977
VERSION
KEYWORDS      HTG.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS      Unpublished
TITLE      Direct Submission
JOURNAL
COMMENT      Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
                  Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
                  3 (bases 1 to 156105)
DOE Joint Genome Institute and Stanford Human Genome Center.
DIRECT SUBMISSION
Submitted (30-NOV-2000) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 156105)
DOE Joint Genome Institute and Stanford Human Genome Center.
DIRECT SUBMISSION
Submitted (06-SEP-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Sep 6, 2001 this sequence version replaced gi:1165049.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center

```


EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

Query Match	0.9%;	Score 21;	DB 3;	Length 156526;
Best Local Similarity	100.0%;	Pred. No. 28;		
Matches 21; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

QY	1508 AAACCAATTCTCAAAATTGAG	1528
Db	100271 AAACCAATTCTCAAAATTGAG	100251

RESULT 84			
AL135926			
LOCUS	158519 bp	DNA	linear
DEFINITION	Human DNA sequence from clone RPL1-37552 on chromosome 1 Contains a		

ACCESSION	AL135926	
VERSION	AL135926.12	GI:9801286
KEYWORDS	HTG.	

REFERENCE

1 (bases 1 to 158519)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

TITLE	COMMENT
Direct Submission Submitted (03-OCT-2000) Sanger Centre, Hinxton, Cambridgeshire CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk On Aug 14, 2000 this sequence version replaced gi:9621473.	

This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems, such as compressions and repeats, but not necessarily within known annotated human repeat sequence elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em; EMBL; Sw; SWISSPROT; Tr; TREMBL; Wb; WORMPEP; information on the WORMPEP database can be found at http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at

RP11-375F2 is from the library RPc11.2 constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://bacpac.med.buffalo.edu/VECTOR:pBACe3.6>

IMPORTANT: This sequence is not the entire insert of clone RP11-375F2. It may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.

The true right end of clone RP11-375F2 is at 156519 in this sequence. The true right end of clone RP1-10C16 is at 100 in this sequence.

FEATURES	Location/Qualifiers
source	1. .158519

```
/organism="Homo sapiens"  
/mol_type="genomic DNA"  
/db_xref="taxon:9606"  
/chromosome="1"
```

```
/chrxxxosome="1"
```

```
/chromosome="1"
```

```
/clone="RP11-375F2"
/clone.lib="RP11-11.2"
repeat_region 460..675
/note="MLTLD repeat: matches 306. .514 of consensus"
repeat_region 1465..1514
/note="25 copies 2 mer to 72% conserved"
repeat_region 1775..1824
/note="MIR repeat: matches 85. .135 of consensus"
repeat_region 3415..3734
/note="L2 repeat: matches 2373. .2707 of consensus"
repeat_region 4294..5519
/note="LIMB7 repeat: matches 4938. .6173 of consensus"
repeat_region 5535..5640
/note="LIMB5 repeat: matches 5322. .5428 of consensus"
repeat_region 5652..5945
/note="AluSq repeat: matches 12. .309 of consensus"
repeat_region 5960..5995
/note="9 copies 4 mer aagg 88% conserved"
repeat_region 5997..6055
/note="LIMB5 repeat: matches 5413. .5471 of consensus"
repeat_region 6061..6158
/note="AluY repeat: matches 214. .311 of consensus"
repeat_region 6160..6855
/note="LIMB5 repeat: matches 5460. .6168 of consensus"
repeat_region 6919..7101
/note="L2 repeat: matches 2206. .2403 of consensus"
repeat_region 7099..7228
/note="L2 repeat: matches 2620. .2750 of consensus"
repeat_region 7828..8069
/note="MER46A repeat: matches 2. .235 of consensus"
repeat_region 8185..8469
/note="AluX repeat: matches 1. .285 of consensus"
repeat_region 8551..8624
/note="MIR repeat: matches 81. .153 of consensus"
misc_feature 8951..9347
/note="match: GSS: Em:B40536"
repeat_region 10287..10464
/note="MIR repeat: matches 28. .215 of consensus"
repeat_region 10588..11321
/note="LIP8 repeat: matches 5371. .6158 of consensus"
repeat_region 11337..11671
/note="MLTLD repeat: matches 13. .381 of consensus"
repeat_region 11672..11971
/note="AluY repeat: matches 1. .295 of consensus"
misc_feature 11855..12218
/note="unidirectional dGTP only"
repeat_region 11972..12090
/note="MLTLD repeat: matches 381. .492 of consensus"
repeat_region 12106..12403
/note="AluX repeat: matches 1. .297 of consensus"
repeat_region 12419..13151
/note="L2 repeat: matches 1683. .2403 of consensus"
misc_feature 13091..13572
/note="match: GSS: Em:AQ887510"
repeat_region 13164..13253
/note="L2 repeat: matches 2576. .2669 of consensus"
repeat_region 13254..13403
/note="75 copies 2 mer at 76% conserved"
repeat_region 13267..13410
/note="24 copies 6 mer tataca 78% conserved"
repeat_region 13654..13742
/note="MLT18 repeat: matches 2. .81 of consensus"
repeat_region 14045..14226
/note="MLT18 repeat: matches 385. .567 of consensus"
misc_feature complement(14529..15003)
/note="match: GSS: Em:AQ798857"
misc_feature 14901..15366
/note="match: GSS: Em:AQ221861"
repeat_region 15371..15453
/note="MIR repeat: matches 60. .153 of consensus"
repeat_region 16536..16633
/note="LIM4 repeat: matches 3093. .3194 of consensus"
repeat_region 16839..17215
/note="LIP47 repeat: matches 5767. .6143 of consensus"
repeat_region 17225..17961
/note="LIP43 repeat: matches 5410. .6146 of consensus"
repeat_region 17962..19671
/note="LIP47 repeat: matches 4077. .5788 of consensus"
repeat_region 19672..20271
/note="LIP47 repeat: matches 3477. .4076 of consensus"
repeat_region 20277..20462
/note="LIM4 repeat: matches 3403. .3585 of consensus"
repeat_region 20479..21023
/note="LIP4a repeat: matches -1540. -.1003 of consensus"
repeat_region 21083..22136
/note="LIP4a repeat: matches -977. .1014 of consensus"
repeat_region 22104..22429
/note="LIM2 repeat: matches 37. .2571 of consensus"
repeat_region 24175..24890
/note="L1 repeat: matches 2118. .2843 of consensus"
repeat_region 24896..25532
/note="LIM1 repeat: matches 5508. .6163 of consensus"
repeat_region 25865..25900
/note="6 copies 6 mer tctgtg 97% conserved"
repeat_region 25903..26184
/note="AluY repeat: matches 1. .282 of consensus"
misc_feature 27110..27647
/note="match: STS: Em:H6045M128"
repeat_region 27233..27346
/note="L2 repeat: matches 2549. .2668 of consensus"
repeat_region 27347..27437
/note="L2 repeat: matches 2064. .2155 of consensus"
repeat_region 27438..27487
/note="LIMC/D repeat: matches 5637. .5685 of consensus"
repeat_region 27552..27933
/note="LIMC4 repeat: matches 6621. .7006 of consensus"
repeat_region 27962..28315
/note="LIMC4 repeat: matches 7094. .7422 of consensus"
repeat_region 28256..28552
/note="LIMC5 repeat: matches 7262. .7589 of consensus"
repeat_region 28629..28938
/note="AluY repeat: matches 1. .310 of consensus"
repeat_region 29045..29218
/note="LIM1 repeat: matches 5338. .5450 of consensus"
repeat_region 30188..30300
/note="MIR repeat: matches 71. .190 of consensus"
repeat_region 30761..30877
/note="L2 repeat: matches 2362. .2474 of consensus"
repeat_region 31253..31494
/note="L2 repeat: matches 466. .712 of consensus"
gene 31722..32025
/gene="DA375F2.1"
/pseudo 31722..32025
/gene="DA375F2.1"
/gene="DA375F2.1" (similar to UBL1 (ubiquitin-like 1 (sestrin)))
match: CDNAS: Em:U72722
match: ESTs: Em:A1191222 Em:A1248769
match: Proteins: Tr:Q9PT08 Tr:Q23759 Sw:P55853 Tr:Q92172
Sw:Q93068 Sw:P55854 Sw:P55855 Sw:Q13351 Sw:Q12306
Tr:Q97102 Tr:O57686"
/pseudo
/codon_start=1
/evidence=not experimental
complement(32129..32630)
/note="match: GSS: Em:AQ747371"
32225..32619
/note="match: GSS: Em:AQ093192"
complement(32404..32630)
/note="match: GSS: Em:AQ195587"
33174..33222
/note="L2 repeat: matches 2624. .2668 of consensus"
33291..33725
```

Query Match

0.9%; Score 21; DB 9; Length 158519;

Best Local Similarity 100.0%; Pred. No. 28;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 12 TGGAGCTGATCACTAGAA 32
Db 123620 TGGAGCTGATCACTAGAA 123640

RESULT 85
AL513307 168544 bp DNA linear PRI 21-OCT-2001
LOCUS Human DNA sequence from clone RP11-239E12 on chromosome 1, complete
DEFINITION sequence.
ACCESSION AL513307
VERSION AL513307.10 GI:16304953
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Moore, M.
TITLE Direct Submission
JOURNAL Submitted (19-OCT-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Oct 21, 2001 this sequence version replaced gi:15022253.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
was generated from part of bacterial clone contigs of human
chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr1
RP11-239E12 is from the library RPCI-11.1 constructed by the group
of Plietier de Jong. For further details see
http://www.choi.org/bacpac/home.htm
VECTOR: pBAC3.6
This sequence is the entire insert of clone RP11-239E12 The true
left end of clone RP11-537P23 is at 23626 in this sequence.
location/Qualifiers
1. 168544
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="1"
/clone="RP11-239E12"
/clone_1id="RPCI-11.1"
91463
91466
misc_feature
91696..91701
/note="Sequence from uni-directional dGTP big dye
terminator reads only."

BASE COUNT 51070 a 31711 c 32256 g 53507 t
ORIGIN

Query Match 0.9%; Score 21; DB 9; Length 168544;
Best Local Similarity 100.0%; Pred. No. 28;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1338 TGTAACTGAGGCTCATTAAG 1358
Db 13658 TGTAACTGAGGCTCATTAAG 13678

RESULT 86
AC010885/c 177385 bp DNA linear PRI 07-NOV-2001
LOCUS Homo sapiens BAC clone RP11-368K23 from 2, complete sequence.
DEFINITION AC010885
ACCESSION AC010885
VERSION AC010885.8 GI:12740528
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS Sulston, J.E. and Waterston, R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE
AUTHORS Pearson, C., Maupin, R. and Haub, K.
TITLE The sequence of Homo sapiens BAC clone RP11-368K23
JOURNAL Unpublished
AUTHORS Waterston, R. H.
TITLE Direct Submission
JOURNAL Submitted (25-SEP-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
AUTHORS Waterston, R. H.
TITLE Direct Submission
JOURNAL Submitted (11-FEB-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
AUTHORS Waterston, R. H.
TITLE Direct Submission
JOURNAL Submitted (09-AUG-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
7 (bases 1 to 177385)
Waterston, R.
REFERENCE
AUTHORS Waterston, R.
TITLE Direct Submission
JOURNAL Submitted (07-NOV-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Feb 10, 2001 this sequence version replaced gi:11968449.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_NH0368K23

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=

30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPc1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E., Tatem, M., Catanesi, J.J. and de Jong, P.J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the right is RP11-328L5, 200 bp overlap.
Actual start of this clone is at base position 1 of RP11-368K23;
actual end is at base position 2422 of RP11-328L5.

FEATURES

source

```

1. 177385
   /organism="Homo sapiens"
   /mol_type="genomic DNA"
   /db_xref="taxon:9606"
   /chromosome="2"
   /map="2"
   /clone="RP11-368K23"
   /clone_1kb="RPc1-11"
   1. 136
   /rpt_family="MALR"
   repeat_region
   137. 581
   /rpt_family="MER1_type"
   repeat_region
   592. 976
   /rpt_family="MER1_type"
   repeat_region
   1310. 1482
   /rpt_family="L2"
   repeat_region
   1985. 2089
   /rpt_family="A-rich"
   repeat_region
   2254. 2334
   /rpt_family="(TCTA)n"
   repeat_region
   2342. 2398
   /rpt_family="U2"
   repeat_region
   2399. 2428
   /rpt_family="(CCTA)n"
   repeat_region
   4645. 4957
   /rpt_family="Alu"
   repeat_region
   5778. 6081
   /rpt_family="Alu"
   repeat_region
   6118. 6238
   /rpt_family="(TCC)n"
   repeat_region
   7895. 8207
   /rpt_family="Alu"
   repeat_region
   9080. 9461
   /rpt_family="MALR"
   repeat_region
   9830. 9853
   /rpt_family="AT-rich"
   repeat_region
   10122. 10237
   /rpt_family="MIR"
   repeat_region
   13454. 14469
   /rpt_family="MER1_type"
   repeat_region
   14518. 16185
   /rpt_family="L1"
   repeat_region
   16906. 17006
   /rpt_family="L2"
   repeat_region
   17543. 17703
   /rpt_family="L2"
   repeat_region
   20062. 20141

```

```

/rpt_family="MIR"
repeat_region
20522. 20835
/rpt_family="MALR"
repeat_region
20836. 21140
/rpt_family="Alu"
repeat_region
21141. 21186
/rpt_family="MALR"
repeat_region
22983. 23329
/rpt_family="L1"
repeat_region
23995. 24023
/rpt_family="AT-rich"
repeat_region
24846. 24908
/rpt_family="(CAG)n"
repeat_region
27752. 27798
/rpt_family="(TG)n"
repeat_region
29252. 29510
/rpt_family="Alu"
repeat_region
29619. 29641
/rpt_family="AT-rich"
repeat_region
30689. 30861
/rpt_family="MER1_type"
repeat_region
32583. 32730
/rpt_family="L2"
repeat_region
34685. 34751
/rpt_family="MIR"
repeat_region
34852. 35051
/rpt_family="MALR"
repeat_region
36518. 36709
/rpt_family="MIR"
repeat_region
37071. 37195
/rpt_family="L2"
repeat_region
37374. 37404
/rpt_family="(TG)n"
repeat_region
37480. 37776
/rpt_family="Alu"
repeat_region
39446. 39624
/rpt_family="MER1_type"
repeat_region
40753. 40954
/rpt_family="MIR"
repeat_region
42387. 42441
/rpt_family="(TCTA)n"
misc_feature
43566. 44026
/note="similar to EST AA424226 (NTD:92103178) zv54g08.r1"
repeat_region
43918. 43999
/rpt_family="MIR"
repeat_region
44393. 45111
/rpt_family="L1"
repeat_region
45116. 46451
/rpt_family="L1"
repeat_region
46091. 46818
/rpt_family="L1"
repeat_region
46932. 47106
/rpt_family="Alu"
repeat_region
47304. 47519
/rpt_family="MIR"
repeat_region
47926. 48269
/rpt_family="L1"
repeat_region
48271. 48357
/rpt_family="L1"
repeat_region
48509. 48591
/rpt_family="(TATG)n"
repeat_region
48647. 48905

```

Query Match 0.9%; Score 21; DB 9; Length 177385;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1606 TGATATATCAGTGAAGG 1626
|||||
DB 41580 TGATATATCAGTGAAGG 41560

RESULT 87

AC104075/c
 LOCUS AC104075 180629 bp DNA linear HTG 03-DEC-2001
 DEFINITION Homo sapiens chromosome UNK clone RP11-371N22, WORKING DRAFT
 SEQUENCE, 44 unordered pieces.
 ACCESSION AC104075.1 GI:17227277
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 REFERENCE 1 (bases 1 to 180629)
 AUTHORS Waterston,R.H.
 TITLE The sequence of Homo sapiens clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 180629)
 AUTHORS Waterston,R.H.
 TITLE Direct Submission
 JOURNAL Submitted (03-DEC-2001) Genome Sequencing Center, Washington
 University School of Medicine, 4444 Forest Park Parkway, St. Louis,
 MO 63108, USA

COMMENT
 ----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 Contact: submissions@wustl.wustl.edu
 ----- Project Information -----
 Center project name: H_NH0371N22
 ----- Summary Statistics -----
 Sequencing vector: M13; 0%
 Sequencing vector: plasmid; 100%
 Chemistry: Dye-primer ET; 0% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 164014 bases at least Q40
 Consensus quality: 167589 bases at least Q30
 Consensus quality: 169890 bases at least Q20

----- NOTE: This is a "working draft" sequence. It currently
 * consists of 44 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1
 1207: contig of 1207 bp in length
 1308: gap of unknown length
 1308: contig of 1926 bp in length
 3333: gap of unknown length
 3334: gap of 1473 bp in length
 4806: contig of 1473 bp in length
 4807: gap of unknown length
 6495: contig of 1589 bp in length
 6496: gap of unknown length
 6595: gap of unknown length
 8313: contig of 1718 bp in length
 8314: gap of unknown length
 8414: gap of 1136 bp in length
 9549: contig of 1136 bp in length
 9550: gap of unknown length
 11447: contig of 1498 bp in length
 11448: gap of unknown length
 11247: gap of 1413 bp in length
 12660: contig of 1413 bp in length
 12661: gap of unknown length
 12760: gap of 1600 bp in length
 14360: contig of 1600 bp in length
 14361: gap of unknown length
 14461: gap of 1388 bp in length
 15848: contig of 1388 bp in length
 15949: gap of unknown length
 17618: contig of 1670 bp in length
 17619: gap of unknown length
 17719: contig of 2090 bp in length
 19808: gap of unknown length
 19809: contig of 2117 bp in length
 19909: contig of 2117 bp in length

22026: gap of unknown length
 22125: gap of 1972 bp in length
 24097: contig of 1972 bp in length
 24197: gap of unknown length
 26002: contig of 1805 bp in length
 26103: gap of unknown length
 28083: contig of 1981 bp in length
 28084: gap of unknown length
 31401: contig of 3218 bp in length
 31402: gap of unknown length
 31502: contig of 3006 bp in length
 34507: gap of unknown length
 34508: gap of 2588 bp in length
 34608: contig of 2588 bp in length
 37195: gap of unknown length
 37295: gap of 3452 bp in length
 40747: contig of 3452 bp in length
 40847: gap of unknown length
 40848: contig of 1698 bp in length
 42545: gap of unknown length
 42646: gap of 3113 bp in length
 42646: contig of 3113 bp in length
 45758: gap of unknown length
 45858: gap of 5086 bp in length
 50944: contig of 5086 bp in length
 51044: gap of unknown length
 51045: contig of 3812 bp in length
 54855: gap of unknown length
 54857: gap of 4548 bp in length
 54957: gap of unknown length
 55904: gap of 4548 bp in length
 55904: gap of unknown length
 59604: gap of 5355 bp in length
 64959: contig of 5355 bp in length
 64960: gap of unknown length
 65059: gap of 3525 bp in length
 68584: contig of 3525 bp in length
 68684: gap of unknown length
 73333: contig of 4649 bp in length
 73334: gap of unknown length
 77984: contig of 4551 bp in length
 78084: gap of unknown length
 83855: contig of 5771 bp in length
 83856: gap of unknown length
 83955: gap of 5946 bp in length
 83956: gap of unknown length
 89901: gap of 5946 bp in length
 89901: gap of unknown length
 90002: contig of 5826 bp in length
 95827: gap of unknown length
 95828: gap of 6348 bp in length
 102275: contig of 6348 bp in length
 102276: gap of unknown length
 102375: gap of 5567 bp in length
 102376: gap of unknown length
 107942: gap of 5567 bp in length
 107942: gap of unknown length
 108042: gap of 7456 bp in length
 108043: gap of unknown length
 115498: contig of 7456 bp in length
 115499: gap of unknown length
 121430: contig of 5832 bp in length
 121431: gap of unknown length
 121530: gap of 10833 bp in length
 121531: gap of unknown length
 132363: contig of 10833 bp in length
 132364: gap of unknown length
 132463: gap of 9795 bp in length
 142258: contig of 9795 bp in length
 142259: gap of unknown length
 142358: gap of 10123 bp in length
 152481: contig of 10123 bp in length
 152482: gap of unknown length
 152581: gap of 9261 bp in length
 152582: gap of unknown length
 161842: contig of 9261 bp in length
 161843: gap of unknown length
 176479: contig of 14537 bp in length
 176479: gap of unknown length
 176579: gap of 1382 bp in length
 176580: gap of unknown length
 177962: contig of 1382 bp in length
 178061: gap of unknown length
 178062: gap of 1197 bp in length
 179258: contig of 1197 bp in length
 179259: gap of unknown length
 180629: contig of 1271 bp in length.

FEATURES
 source
 1.180629
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="UNK"
 /clone="RP11-371N22"
 1.1207
 /note="assembly_name:Contig10"
 1308...3333
 /note="assembly_name:Contig13"

misc_feature
 misc_feature
 misc_feature

	misc_feature	3334..4806	/note="assembly_name:Contig14"
	misc_feature	4907..6495	/note="assembly_name:Contig15"
	misc_feature	6596..8313	/note="assembly_name:Contig16"
	misc_feature	8414..9549	/note="assembly_name:Contig17"
	misc_feature	9650..11147	/note="assembly_name:Contig18"
	misc_feature	11248..12660	/note="assembly_name:Contig19"
	misc_feature	12761..14360	/note="assembly_name:Contig20"
	misc_feature	14461..15848	/note="assembly_name:Contig21"
	misc_feature	15949..17618	/note="assembly_name:Contig22"
	misc_feature	17719..19808	/note="assembly_name:Contig23"
	misc_feature	19909..22025	/note="assembly_name:Contig24"
	misc_feature	22126..24097	/note="assembly_name:Contig25"
	misc_feature	24198..26002	/note="assembly_name:Contig26"
	misc_feature	26103..28083	/note="assembly_name:Contig27"
	misc_feature	28184..31401	/note="assembly_name:Contig28"
	misc_feature	31502..34507	/note="assembly_name:Contig29"
	misc_feature	34608..37195	/note="assembly_name:Contig30"
	misc_feature	37296..40747	/note="assembly_name:Contig31"
	misc_feature	40848..42545	/note="assembly_name:Contig32"
	misc_feature	42646..45758	/note="assembly_name:Contig33"
	misc_feature	45859..50944	/note="assembly_name:Contig34"
	misc_feature	51045..54856	/note="assembly_name:Contig35"
	misc_feature	54957..59504	/note="assembly_name:Contig36"
	misc_feature	59605..64959	/note="assembly_name:Contig37"
	misc_feature	65060..68584	/note="assembly_name:Contig38"
	misc_feature	68685..73333	/note="assembly_name:Contig39"
	misc_feature	73434..77984	/note="assembly_name:Contig40"
	misc_feature	78085..83855	

Query Match 0.9% ; Score 21; DB 2; Length 180629;
Best Local Similarity 100.0%; Pred.No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Oy      1606 TGATTAATCAGTGAAGAAGG 1626  
        |||||  
Db       135343 TGAATAATCACTGATGAAGG 135323
```

RESULT 88
 LOCUS AC117257:c 180643 bp DNA linear HTG_09-MAY-2002
 DEFINITION Mus musculus chromosome UNK clone RP24-484P21, WORKING DRAFT
 AC117257
 SEQUENCE, 5 unordered pieces.
 AC117257.2 GI:20514968
 KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.

```

SOURCE
Mus musculus (house mouse)

REFERENCE
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 180643)
McPherson,J.D. and Waterston,R.H.
The sequence of Mus musculus clone
Unpublished
2 (bases 1 to 180643)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (08-APR-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 180643)
McPherson,J.D. and Waterston,R.H.
Direct Submission
Submitted (09-MAY-2002) Genome Sequencing Center, 4444 Forest Park
Parkway, St. Louis, MO 63108, USA
On May 9, 2002 this sequence version replaced gi:20069845.

COMMENT
----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site:http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M_BB0484F21
----- Summary Statistics -----
Sequencing vector: M13, 0%
Sequencing vector: plasmid, 100%
Chemistry: Dye-primer ET; 0% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 178168 bases at least Q40
Consensus quality: 178671 bases at least Q40
Consensus quality: 179048 bases at least Q20
Insert size: 193000; agarose-fp
Insert size: 180243; sum-of-contigs
Quality coverage: 9.47 in Q20 bases; agarose-fp
Quality coverage: 10.41 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1
63632: contig of 63632 bp in length
*
* 63633 63732: gap of unknown length
*
* 63733 160534: contig of 96802 bp in length
*
* 160535 160634: gap of unknown length
*
* 160635 162109: contig of 1475 bp in length
*
* 162110 162209: gap of unknown length
*
* 162210 171033: contig of 8824 bp in length
*
* 171034 171133: gap of unknown length
*
* 171134 180643: contig of 9510 bp in length.

Location/Qualifiers
1..180643
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-484F21"
1..63632
/note="assembly_name:Contig10"
63733..160534
/note="assembly_name:Contig11"
160635..162109
/note="assembly_name:Contig7"
162210..171033

```

```

/note="assembly_name:Contig9"
misc_feature
171134..180643
/note="assembly_name:Contig9"
BASE COUNT 50723 a 37481 c 38933 g 53097 t 409 others
ORIGIN

Query Match 0.9%; Score 21; DB 2; Length 180643;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1212 CATCGAAGCTTTAAATC 1232
Db 84507 CATCGAAGCTTTAAATC 84487

RESULT 89
AC108815/c 181547 bp DNA linear HTG 08-MAR-2003
LOCUS Mus musculus clone RP23-239013, WORKING DRAFT SEQUENCE, 5 unordered
DEFINITION pieces.
AC108815
AC108815.4 GI:28882391
VERSION HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 181547)
Birten,B., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-239013
Unpublished
2 (bases 1 to 181547)
Birten,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barina,N., Bastien,V., Boguslavsky,L., Boukhgalter,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Cooke,P., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cokep,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Fair,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatae,A., Kelle,C., Laroque,K., Lamatares,R.,
Lander,T., Lehoczy,J., Levine,R., Liu,G., MacLean,C.,
Macdonald,P., Major,J., Margulis,N., Matthews,C., McCarthy,M.,
Mcwan,P., McKernan,K., Meldrum,J., Mensu,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Pollara,V., Rogov,P., Roman,S.,
Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P., Roman,J.,
Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuback,R., Seaman,S.,
Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Testfay,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Triggilio,J., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zemdek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (31-JAN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 181547)
Birten,B., Nusbaum,C., Lander,E., Aboueleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barina,N., Bastien,V., Bloom,T.,
Boguslavsky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatae,A., Kelle,C., Lander,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
Maddison,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrum,J., Mensu,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,

```

```

TITLE
JOURNAL
COMMENT
Roman,J., Schauer,S., Schuback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Testfay,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zemdek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (08-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Mar 8, 2003 this sequence version replaced gi:28195423.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
-----
Genome Center
Center code: WIBR
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
-----
Project Information
Center project name: L20929
Center clone name: 239 O.13
-----
Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 180713 bases at least Q40
Consensus quality: 180940 bases at least Q30
Consensus quality: 181059 bases at least Q20
Insert size: 172000; agarose-fp
Insert size: 181147; sum-of-contigs
Quality coverage: 8.0 in Q20 bases; agarose-fp
Quality coverage: 7.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 15598: contig of 15598 bp in length
* 15599 15698: gap of 100 bp
* 15699 17443: contig of 1745 bp in length
* 17444 17543: gap of 100 bp
* 17544 39808: contig of 2265 bp in length
* 39809 39908: gap of 100 bp
* 39909 76850: contig of 36942 bp in length
* 76851 76950: gap of 100 bp
* 76951 181547: contig of 104597 bp in length.
Location/Qualifiers
1. 181547
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/clone_id="RP23-239013"
/clone_lib="RP23-239013"
/clone_lib="RP23-239013"
1. 15598
/note="assembly_fragment"
clone_end:SP6
vector_side:left"
15699..17443
/note="assembly_fragment"
17544..39808
/note="assembly_fragment"
39909..76850
/note="assembly_fragment"
76951..181547
/note="assembly_fragment"
clone_end:T7
vector_side:right"
BASE COUNT 52438 a 35684 c 36618 g 56407 t 400 others
ORIGIN

```

Query Match	0.9%	Score 21	DB 2	Length 181547
Best Local Similarity	100.0%	Pred. No. 27		
Matches	21	Conservative	0	Mismatches 0; Indels 0; Gaps 0
OY	1539	ACGAAGTCGATTCCTCAAT	1559	
DB	113083	ACGAAGTCGATTCCTCAAT	113063	

RESULT	90				
LOCUS	HS109M15/c				
DEFINITION	Homo sapiens chromosome 9 BAC Rpl1-109M15,	184365 bp DNA	linear	PRI 14-FEB-2003	
ACCESSION	AL513424 AL353775			complete sequence.	
VERSION	AL513424.3	G1:28400938			
KEYWORDS	HTG.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. 1 (bases 1 to 184365) Scharfe,M., Conrad,A., Hornischer,K., Loehnert,T.H., Thies,S. and Bloecher,H. Direct Submission Submitted (09-SEP-2000) GBF, Dept. of Genome Analysis, Mascheroder Weg 1, D-38124 Braunschweig, Germany, E-mail: info.genome@gbf.de On Feb 17, 2003 this sequence version replaced g1:15384820.				
TITLE	JOURNAL	AUTHORS	COMMENT		

```

http://webcbe.sanger.ac.uk/cgi-bin/display?db=accedb9&grep=109M15
----- Genome Center
Center: GBF, Braunschweig
Center code: GBF
Web site: http://genome.gbf.de/
Contact: info.genome@gbf.de
----- Project Information
Center project name:
Center Clone name: 109M15
----- Summary Statistics
Sequencing vector: ##;
Chemistry: Dye-terminator-BigDye: 45% of reads
Chemistry: Dye-terminator-amersham: 47% of reads
Chemistry: Dye-primer-amersham: 8% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 0 bases at least Q40
Consensus quality: 0 bases at least Q30
Consensus quality: 0 bases at least Q20
Estimated insert size: ##; agarose-tp estimation
Estimated insert size: 184365; sum-of-ctnigs estimation
-----
PROGRAMS AND PARAMETERS USED FOR ANNOTATION:
+++++
+   Analysis and annotation were performed with the automatic +
+   'first-pass' annotation and submission tool +
+   'AnnotMlter' (Hornischer & Bloeker). +
+   Programs used by 'AnnotMlter': +
+++++
+ GeneFinder (Green), Vers. 084 +
+ Organism: human +
+ Genscan (Burge & Karlin), Vers. 1.0 +
+   Used matrix: vertebrate; Minimum score: 0 +
+   Grail (Xu et al.), Vers. 1.3 +
+   Organism: human +
+   Mzef (Zhang) +
+   Prior probability: 0.04; Overlapping number: 0 > Xpound (Thomas +
+   & Skolnick) +
+   Base score cutoff: 0.2; Minimal exon length: 3 bp > 'Repeats': +
+   BLASTN 2.0.14 (Altschul et al.) +
+   Database(s): * RepBase: ALU (human), released 22-DEC-1995 .

```

```

FEATURES
source
1. 184365

* Repbase: THR (human), released 22-DEC-1995 .
* Repbase: LI (primate), released 22-DEC-1995 .
* Repbase: MER (primate), released 22-DEC-1995 .
* Repbase: MIR (primate), released 22-DEC-1995 .
* Repbase: THE (primate), released 22-DEC-1995 .
* Minimum identity: 70 %; Minimum score: 60;
> 'ESTs': BLASTN 2.0.14 (Altschul et al.)
* Database(s): * embd (EST, human), released -DEC-
* emb1 (EST, other), released -DEC-
(EST), released -DEC-
* Using sequence with masked repeats
* Minimum score: 60; Minimum identity: 70 %;
> 'Tandem Repeats': GDE 2.2 option 'tandem'
* Minimum length 2 bp; Maximum length 20 bp; Score threshold 20
* Treat N's as mismatches? YES; Allow uniform consensus? NO >
> 'Inverted Repeats': GDE 2.2 option 'inverted'
> 'Micro Satellites': GDE 2.2 option 'spunuk' (Abajlan) > 'CPG
Islands': GDE 2.2 option 'cpgr'
* CPG island region size 100 bp;
* Minimum GC contents 50 %; Observed/Expected 0.6 > 'STS Scan':
e-PCR (Schuler)
* Margin: 50; Number of mismatches allowed: 0; Word size: 7 .
* SWS database: 'dbSTS markers'
> 'tRNA Scan': tRNAscan-SE (Lowe & Eddy), Vers. 1.11.
Location/Qualifiers
1. 184365

```

```

misc_feature
1. .184365
/note="assembly_fragment-clone.end:T7-vector.side:left
assembly_fragment-clone.end:Sp6-vector.side:right"
64. .145
/note="91% identity: matches 62. .141 of consensus"
/rpt_family="THE"
114. .224
/note="94% identity: matches 85. .195 of consensus"
/rpt_family="THE"
1883. .1962
/note="XPOUND prediction, score = 0.375"
complement(1943. .1981)
/note="XPOUND prediction, score = 0.240"
complement(2006. .2017)
/note="XPOUND prediction, score = 0.212"
complement(2323. .2378)
/note="91% identity: matches 369. .424 of consensus"
/rpt_family="ALU"
complement(2371. .2457)
/note="86% identity: matches 57. .143 of consensus"
/rpt_family="ALU"
complement(2475. .2583)
/note="88% identity: matches 334. .442 of consensus"
/rpt_family="ALU"
2479. .2666
/note="MZEF prediction, score = 0.982"
2487. .2582
/note="85% identity: matches 829. .924 of consensus"
/rpt_family="L1"
complement(2696. .2752)
/note="XPOUND prediction, score = 0.274"
complement(3150. .3200)
/note="92% identity: matches 61. .111 of consensus"
/rpt_family="THE"
5882. .5901
/note="GA repeat"
6165. .6202
/note="XPOUND prediction, score = 0.334"
complement(8738. .8890)
/note="GENSCAN prediction, score = 6.37"

```

```

exon          9503. .9589
              /note="XPOUND prediction, score = 0.273"
exon          10081. .10110
              /note="XPOUND prediction, score = 0.203"
satellite     10686. 10698
              /note="TCTT repeat"
exon          10724. .10894
              /note="MZF prediction, score = 0.808"
exon          11716. .11793
              /note="XPOUND prediction, score = 0.631"
exon          12302. .12394
              /note="XPOUND prediction, score = 0.626"
satellite     12705. .12719
              /note="TTTA repeat"
repeat_region 12709. .13002
              /note="IR1 79% complementary to IR1' (13162. .13453)"
              /rpt_type=INVERTED
repeat_region 12721. .13003
              /note="82% identity: matches 312. .597 of consensus"
              /rpt_family="LI"
repeat_region 12722. .12984
              /note="88% identity: matches 154. .416 of consensus"
              /rpt_family="ALU"
repeat_region 13162. .13453
              /note="IR1' 79% complementary to IR1 (12709. .13002)"
              /rpt_type=INVERTED
repeat_region 13171. .13412
              /note="90% identity: matches 8. .249 of consensus"
              /rpt_family="ALU"
repeat_region 13176. .13440
              /note="84% identity: matches 162. .424 of consensus"
              /rpt_family="LI"
exon          14673. .14694
              /note="XPOUND prediction, score = 0.227"
exon          14887. .14942
              /note="XPOUND prediction, score = 0.640"
              /note="84% identity: matches 15894. .15983"
              /rpt_family="THE"
repeat_region 15971. .16005
              /note="XPOUND prediction, score = 0.319"
exon          16092. .16146
              /note="XPOUND prediction, score = 0.244"
exon          16224. .16224
              /note="GENSCAN prediction, score = 0.90"
exon          17338. .17364
              /note="XPOUND prediction, score = 0.382"
exon          18420. .18547
              /note="XPOUND prediction, score = 0.422"
exon          18859. .18883
              /note="XPOUND prediction, score = 0.325"
repeat_region 19806. .19835
              /note="homology = 93.3%, counts = 15"
              /rpt_family="ac repeat"
              /rpt_type=TANDEM
satellite     19811. .19835
              /note="CA repeat"
              /note="complement(20400. .20508)"
              /note="88% identity: matches 494. .602 of consensus"
              /rpt_family="ALU"
repeat_region 20400. .20484
              /note="87% identity: matches 175. .259 of consensus"
              /rpt_family="LI"
exon          21590. .21709
              /note="MZF prediction, score = 0.739"

```

```

Query Match      0.9% Score 21; DB 9; Length 184365;
Best Local Similarity 100.0% Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

OY 560 CCTCATGAGATCATTCACAT 580
    |||||
DB 131311 CCTCATGAGATCATTCACAT 131291

```

```

RESULT 91
LOCUS AC120080
DEFINITION Rattus norvegicus clone CH230-373p11. *** SEQUENCING IN PROGRESS
ACCESSION AC120080
VERSION AC120080.5 GI:25094221
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
AUTHORS 1 (bases 1 to 188307)
Muzny,D.,Marie, Metzker,M.,Lee, Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooke,S., Amin,A., Angiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Duthin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Fall,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Friser,C.M., Gablis,A., Gante,R., Garcia,A., Garner,T., Garza,M.,
Georgiev,E., Geer,K., Gill,R., Grady,M., Guerra,M., Guevara,W.,
Gunnarsson,P., Healand,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hognes,M.,
Hollins,B., Howlis,S., Huily,S., Hume,J., Idebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Kapatthy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowals,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenzowa,L., Louissegh,H., Lozano,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindaratne,M., Mahmood,M., Malloy,K., Mangum,A.,
Mangum,B., Mapue,P., Martin,K., Martin,R., Martinez,E.,
Mawhinney,S., McLeod,M.P., McNeill,T.Z., Meener,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Mundasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokemele,O., Okwuonu,G., Olarinmusaogun,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Polidexter,A., Popovic,D., Primus,E., Pu,L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojars,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Sma's,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Swack,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villaseña,D., Waldron,L., Walker,B., Wang,D.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczky,R., Woodson,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 188307)
Worley,K.C.
Direct Submission
Submitted (03-MAY-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

```

REFERENCE 3 (bases 1 to 188307)
 AUTHORS Rat Genome Sequencing Consortium.
 TITLE Direct Submission
 JOURNAL Submitted (19-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Nov 19, 2002 this sequence version replaced gi:23194715. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center Project name: GVTD

Center Clone name: CH230-373P11

----- Summary Statistics

Assembly program: PHRAP; version 0.990329

Consensus quality: 176123 bases at least Q40

Consensus quality: 178360 bases at least Q30

Consensus quality: 179208 bases at least Q20

Estimated insert size: 182713; sum-of-contigs estimation

Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length

(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).

* NOTE: This is a 'working draft' sequence. It currently

* consists of 2 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 186944: contig of 186944 bp in length

* 186945 187044: gap of unknown length

* 187045 188307: contig of 1263 bp in length.

FEATURES
 source 1.188307
 Location/Qualifiers
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-373P11"

misc_feature

1.2047

/note="wgs end_extension"

clone_end:77"

6257..7382.

/note="wgs end_extension"

clone_end:77"

14936..15763

/note="clone_boundary"

clone_end:77

site:

end_sequence:BZ241333"

185598..186478

/note="clone boundary"

clone_end:Sp6

site:

end_sequence:BZ241334"

BASE COUNT 63393 a 35575 c 34850 g 46244 t 8245 others
 ORIGIN

Query Match 0.9%; Score 21; DB 2; Length 188307;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

cy 1592 ATTTTGTGATGACATGATATA 1612
 |||||
 db 14495 ATTTTGTGATGACATGATATA 14515

RESULT 92

BX537317/c

LOCUS

DEFINITION

uncloned pieces.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Danio rerio

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Actinopterygii; Neopterygii; Teleostei; Osteichthyes;

Cypriniformes; Cyprinidae; Danio.

1 (bases 1 to 191750)

REFERENCE

AUTHORS

TITLE

JOURNAL

----- Summary Statistics

Assembly program: XGAP4; version 4.5

Chemistry: Dye-terminator; 100% of reads

Consensus quality: 190069 bases at least Q40

Consensus quality: 190312 bases at least Q30

Consensus quality: 190418 bases at least Q20

Insert size: 191350; sum-of-contigs

Insert size: 180369; 6.9% error; agarose-fp

Quality coverage: 6.99x in Q20 bases; sum-of-contigs Quality

coverage: 7.42x in Q20 bases; agarose-fp

* NOTE: This is a 'working draft' sequence. It currently

* consists of 5 contigs. The true order of the pieces

* is not known and their order in this sequence record is

* arbitrary. Gaps between the contigs are represented as

* runs of N, but the exact sizes of the gaps are unknown.

* This record will be updated with the finished sequence

* as soon as it is available and the accession number will

* be preserved.

* 1 57919: contig of 57919 bp in length

* 57920 58019: gap of 100 bp

* 58020 64419: contig of 6400 bp in length

* 64420 64519: gap of 100 bp

* 64520 79865: contig of 15346 bp in length

* 79866 79965: gap of 100 bp

* 79966 138076: contig of 58111 bp in length

* 138077 138176: gap of 100 bp

* 138177 191750: contig of 53574 bp in length.

Location/Qualifiers

1.191750

/organism="Danio rerio"

/mol_type="genomic DNA"

/db_xref="taxon:7955"

/clone="DKEX-148LA"

/clone_lib="Daniokey"

1..57919

/note="assembly fragment:00083"

fragment_chain:1"

misc_feature

1..57919

```

misc_feature      58020..64419
                  /note="assembly_fragment:00193
                  fragment_chain:1"
                  64520..79865
misc_feature      /note="assembly_fragment:00994
                  fragment_chain:2"
                  79966..138076
misc_feature      /note="assembly_fragment:00346
                  fragment_chain:2"
                  138177..191750
BASE COUNT      61624 a 34551 c 34190 g 60700 t      685 others
ORIGIN
Query Match      0.9%; Score 21; DB 2; Length 191750;
Best Local Similarity 100.0%; Pred.No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0.

Oy      2085 ACAGTGAAGATCCAAATAGA 2105
        |||||
Db      103553 ACAGTGAAGATCCAAATAGA 103533

RESULT_93
LOCUS      AP001523/3
DEFINITION Homo sapiens chromosome 11 clone RP11-701D3 map 11q14, WORKING
           DRAFT SEQUENCE, 27 unordered pieces.
ACCESSION      AP001523
VERSION      AP001523.2 GI:8117364
KEYWORDS      HTG; PHASE1; HTGS_DRAT.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE      1 (bases 1 to 193433)
AUTHORS      Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
              Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
              Homo sapiens 193,433 genomic DNA of 11q14
              Published Only in Database (2000)
              2 (bases 1 to 193433)
              Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
              Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
              Direct Submission
              Submitted (22-MAR-2000) Masahira Hattori, The Institute of Physical
              and Chemical Research (RIKEN), Genomic Sciences Center (GSC),
              Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
              Japan (E-mail:hattori@gsc.riken.go.jp,
              URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
              Fax:81-42-778-9924)
              On May 30, 2000 this sequence version replaced gi:7328537.

COMMENT
        ----- Genome Center
        Center: RIKEN Genomic Sciences Center (GSC)
        Center code: RIKEN
        Web site: http://hgp.gsc.riken.go.jp/
        Contact: hattori@gsc.riken.go.jp
        ----- Project Information
        Center project name: HumDrafc11
        Center clone name: RP11-701D3
        ----- Summary Statistics
        Sequencing vector: PCR products, 100% of reads
        Chemistry: Dye-terminator ET-amersham; 100% of reads
        Assembly program: Phrap; version 0.990329
        Consensus quality: 177276 bases at least Q40
        Consensus quality: 185320 bases at least Q30
        Consensus quality: 188804 bases at least Q20
        Insert size: 190833; sum-of-contigs
        Quality coverage: 4.92x in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of
27 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps

```

are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```
1      18117: contig of 18117 bp in length
18818 36734: contig of 17917 bp in length
36835 46657: contig of 9823 bp in length
46758 60859: contig of 14102 bp in length
60960 72573: contig of 11714 bp in length
72774 85821: contig of 13048 bp in length
85922 95147: contig of 9226 bp in length
95248 106548: contig of 11301 bp in length
106649 116133: contig of 9485 bp in length
116234 125686: contig of 9453 bp in length
125787 131853: contig of 6667 bp in length
131854 131953: gap of 100 bp
131954 137782: contig of 5825 bp in length
137883 137882: gap of 100 bp
144531 144631: contig of 6649 bp in length
144632 144631: gap of 100 bp
144632 148906: contig of 4275 bp in length
148907 149006: gap of 100 bp
149007 154806: contig of 5800 bp in length
154807 154906: gap of 100 bp
154907 160956: contig of 6050 bp in length
160957 161056: gap of 100 bp
161057 165055: contig of 3999 bp in length
165056 165155: gap of 100 bp
```

Sequence updated (26-May-2000).

NOTE: This is a 'working draft' sequence. It currently consists of 27 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

```

* 165156 170635: contig of 5480 bp in length
* 170636 170735: gap of 100 bp
* 170736 175232: contig of 4497 bp in length
* 175233 175332: gap of 100 bp
* 175333 178109: contig of 2777 bp in length
* 178110 178209: gap of 100 bp
* 178210 181915: contig of 3706 bp in length
* 182015 182015: gap of 100 bp
* 182016 183596: contig of 1581 bp in length
* 183597 183696: gap of 100 bp
* 183697 186393: contig of 2697 bp in length
* 186394 186493: gap of 100 bp
* 186494 188330: contig of 1837 bp in length
* 188331 188430: gap of 100 bp
* 188431 190127: contig of 1697 bp in length
* 190128 190227: gap of 100 bp
* 190228 191605: contig of 1378 bp in length
* 191606 191705: gap of 100 bp
* 191706 193433: contig of 1728 bp in length.
FEATURES
source 1..193433
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="11"
        /map="11q14"
        /clone="RP11-701D3"
1..18717
misc_feature /note="assembly_fragment clone_end:SP6 vector_side:right"
misc_feature 18818..36734
            /note="assembly_fragment"
misc_feature 36835..46657
            /note="assembly_fragment clone_end:T7 vector_side:right"
46758..60859
misc_feature /note="assembly_fragment"
60960..72673
misc_feature /note="assembly_fragment"
72774..85821
misc_feature /note="assembly_fragment"
85922..95147
misc_feature /note="assembly_fragment"
95248..106548
misc_feature /note="assembly_fragment"
106649..116133
misc_feature /note="assembly_fragment"
116234..125686
misc_feature /note="assembly_fragment"
125787..131853
misc_feature /note="assembly_fragment"
131954..137782
misc_feature /note="assembly_fragment"
137883..144531
misc_feature /note="assembly_fragment"
144632..148906
misc_feature /note="assembly_fragment"
149007..154806
misc_feature /note="assembly_fragment"
154907..160936
misc_feature /note="assembly_fragment"
161057..165055
misc_feature /note="assembly_fragment"
165156..170635
misc_feature /note="assembly_fragment"
170736..175232
misc_feature /note="assembly_fragment"
175333..178109
misc_feature /note="assembly_fragment"
178210..181915
misc_feature /note="assembly_fragment"
182016..183596
misc_feature /note="assembly_fragment"
183697..186393
            /note="assembly_fragment"

```

```

misc_feature 186494..188330
            /note="assembly_fragment"
misc_feature 188431..190127
            /note="assembly_fragment"
misc_feature 190228..191605
            /note="assembly_fragment"
Query Match 0.9%; Score 21; DB 2; Length 193433;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2319 CAGAGCAGAAAAATTACTTG 2319
    |||||
Db 13274 CAGAGCAGAAAAATTACTTG 13254
RESULT 94
AC024569 193547 bp DNA linear PRI 12-DEC-2001
LOCUS Homo sapiens chromosome 5 clone CTD-2188H20, complete sequence.
AC024569
ACCESSION AC024569.5 GI:17530706
VERSION HTG.
KEYWORDS SOURCE
ORGANISM Homo sapiens (human)
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
1 (bases 1 to 193547)
DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Unpublished
2 (bases 1 to 193547)
DOE Joint Genome Institute.
JOURNAL Direct Submission
Submitted (29-FEB-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
3 (bases 1 to 193547)
DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
Submitted (23-OCT-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 193547)
DOE Joint Genome Institute and Stanford Human Genome Center.
JOURNAL Direct Submission
Submitted (12-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
On Dec 12, 2001 this sequence version replaced gi:16328262.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.6% of Sequence;
Estimated Total Number of Errors is 0.4.
FEATURES
source 1..193547
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="5"
        /clone="CTD-2188H20"
BASE COUNT 63355 a 36242 c 35229 g 58721 t
ORIGIN
Query Match 0.9%; Score 21; DB 9; Length 193547;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1227 AAAATCTGAAAGCAAGTTAT 1247
    |||||
Db 167060 AAAATCTGAAAGCAAGTTAT 167040
RESULT 95

```

AC114491 207698 bp DNA linear PRI 30-APR-2002
 LOCUS Homo sapiens chromosome 1 clone RP11-270C12, complete sequence.
 DEFINITION AC114491 AL160002
 ACCESSION AC114491.1 GI:19310310
 VERSION
 KEYWORDS HTG.
 ORGANISM Homo sapiens (human)
 SOURCE
 REFERENCE
 AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 207698)
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
 Saenphimachak, C., Phelps, K.A., Buckley, D., Raymond, C. and
 Haugen, E.D.
 TITLE Direct Submission
 JOURNAL
 REFERENCE 2 (bases 1 to 207698)
 AUTHORS Kaul, R.K., Olson, M.V., Raymond, C. and Haugen, E.D.
 TITLE Direct Submission
 JOURNAL Submitted (09-MAR-2002) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 3 (bases 1 to 207698)
 Kaul, R.K., Olson, M.V., Zhou, Y., James, R.A., Rouse, G., Wu, Z.,
 Saenphimachak, C., Phelps, K.A., Buckley, D., Raymond, C. and
 Haugen, E.D.
 TITLE Direct Submission
 JOURNAL Submitted (30-APR-2002) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 On Mar 9, 2002 this sequence version replaced gi:9212681.
 COMMENT
 ----- Genome Center
 Center: University of Washington Genome Center
 Center Code: UMGC
 Web site: http://www.genome.washington.edu
 Contact: uwgctgs@u.washington.edu
 Drafting Center: SC
 ----- Project Information
 Center project name: chr-1
 Center clone name: RP11-270C12 (sc0661)
 ----- Summary Statistics
 Sequencing vector: plasmid; 43% of reads
 Chemistry: Dye-terminator ET; 89% of reads
 Chemistry: Dye-terminator Big Dye; 11% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 207698 bases at least Q40
 Consensus quality: 207698 bases at least Q30
 Consensus quality: 207698 bases at least Q20
 Insert size: 207698; sum-of-contigs
 Quality coverage: 9.9x in Q20 bases; sum-of-contigs

 Overlapping Sequences:
 5': Mapping in progress
 3': RP11-436H6 AL513187, 2000-bp overlap

 Sequence Quality Assessment:
 This entry has been annotated with sequence quality
 estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than
 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the
 GenBank flat file format but are available as part
 of this entry's ASN.1 file.

 This sequence was finished as follows unless otherwise noted:
 all regions were either double-stranded or sequenced with an
 alternate chemistry or covered by high quality data (i.e., Phred
 quality >= 30); an attempt was made to resolve all sequencing
 problems, such as compressions and repeats; all regions were
 covered by at least one plasmid subclone or more than one M13
 subclone; and the assembly was confirmed by restriction digest.

 Sequence Validation:

This sequence has been validated by Multiple Complete Digest
 fingerprinting. Comparison of the experimentally derived digest
 fragments with sequence-predicted fragments is given below.
 The electronically-digested sequence consists of both insert and
 vector, in order to accurately represent the entire circular BAC.
 Small fragments below a variable cutoff (approximately 400-800 bp)
 are not resolved in the fingerprint and hence do not appear
 in the table. There are no significant remaining discrepancies
 between the experimental and predicted values. Uniquely ordered
 fragments are separated by dashed lines.

ECORI				BglII				HindIII			
SeqDerMap	Fingerprint	SeqDerMap	Fingerprint	SeqDerMap	Fingerprint	SeqDerMap	Fingerprint	SeqDerMap	Fingerprint	SeqDerMap	Fingerprint
8696	8845	3787	3787	3787	3784						
6	<800	2067	2058	6382	6382						
588	<800	19478	19218	512	<800						
3748	3660	7236	7270	449	<800						
4201	4075	1838	1836	2164	2129						
744	<800	7859	7854	505	<800						
2503	2630	23702	24165	5095	4963						
295	<800	9	<800	1569	1501						
2266	2381	8399	8565	9462	9221						
6491	6358	2729	2720	412	<800						
5780	5747	5829	5843	733	731						
2860	2878	4831	4892	5942	6003						
9091	8845	1250	1249	8894	8677						
1164	1111	8593	8565	3476	3591						
568	<800	2453	2554	6483	6382						
2085	2087	9460	9578	9270	9221						
5801	5747	2951	3018	2437	2491						
1934	1908	6477	6410	3574	3591						
164	<800	1586	1556	4517	4448						
1943	1908	1275	1249	3213	3434						
939	926	1215	1249	2869	2693						
10827	10607	1710	1692	7440	7392						
2080	2087	14894	14529	2683	2693						
5992	5747	3207	3215	3127	3064						
4371	4397	5525	5497	5845	5812						
3178	3165	1316	1249	520	<800						
6387	6358	6301	6410	238	<800						
141	<800	5423	5497	1614	1620						
6250	6358	248	<800	2527	2491						

557	<800	1318	1249	11716	11413
3000	3024	2673	2720	4827	4714
4453	4397	4615	4535	553	<800
22641	22765	9793	9578	16194	16511
8253	8138	6907	6913	331	<800
845	839	3005	3018	11097	10703
1153	1111	11034	10774	728	731
664	<800	2535	2554	2704	2827
1642	1578	3438	3410	11091	10703
2912	2878	2551	2554	10650	10274
1519	1511	4505	4535	5094	4863
5483	5369	2375	2554	179	<800
2134	2087			3900	3784
7431	7350			950	954
2127	2087			250	<800
3734	3660			3846	3784
1552	1511			3208	3194
5752	5747			1537	1501
2499	2506			632	<800
2447	2506			716	<800
1476	1511			2097	2048
1553	1511			41	<800
678	<800			3632	3784
3346	3365			336	<800
540	<800			15	<800
2422	2381			2246	2232
6476	6358			6215	6382

Query Match 0.9% Score 21; DB 9; Length 207698;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1219 AACTGTTAAATCTGGAAG 1239
 Db 9914 AACTGTTAAATCTGGAAG 9934

RESULT 96
 AC119573 219687 bp DNA 1linear HTG 23-APR-2003
 LOCUS AC119573
 DEFINITION Mus musculus clone RP23-75H23, WORKING DRAFT SEQUENCE, 5 ordered
 pieces.
 AC119573
 AC119573 GI:30017949
 VERSION AC119573.3
 KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT.

SOURCE
 ORGANISM
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 219687)
 Birren,B., Nussbaum,C. and Lander,E.
 TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 2 (bases 1 to 219687)
 Birren,B., Linton,L., Nussbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
 Bouckgalter,B., Brown,A., Camarata,J., Campoliano,A., Chang,J.,
 Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
 Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Faro,S., Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Lacroque,K., Lamazares,R.,
 Landers,T., Lehoczyk,J., Levine,R., Lindblad-Toh,K., Liu,G.,
 Maclean,C., Macdonald,P., Major,J., Margulis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., Meldrum,J., Meneus,L.,
 Mihova,T., Mienga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunhkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupbach,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Testaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Triggillo,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (28-APR-2002) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 3 (bases 1 to 219687)
 REFERENCE
 JOURNAL
 TITLE
 AUTHORS
 Birren,B., Nussbaum,C., Lander,E., Abouelell,A., Allen,N.,
 Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
 Boguslavskiy,L., Bouckgalter,B., Camarata,J., Chang,J., Choepel,Y.,
 Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,
 Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
 Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J., Gardyna,S.,
 Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
 Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
 Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
 Macdonald,P., Major,J., Manning,D., Matthews,C., McCarthy,M.,
 Meldrum,J., Meneus,L., Mihova,T., Mienga,V., Murphy,T., Naylor,J.,
 Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,J., Peterson,K., Phunhkhang,P., Pierre,N.,
 Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
 Roman,J., Schauer,S., Schupbach,R., Seaman,S., Severy,P., Smith,C.,
 Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
 Talamas,J., Testaye,S., Theodore,J., Topham,K., Travers,M.,
 Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (23-APR-2003) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Apr 17, 2003 this sequence version replaced g1:20389373.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence submissions@genome.wi.mit.edu
 Project Information
 Center project name: 75_H23
 Center clone name: 75_H23

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 5 contigs. Gaps between the contigs

```

* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 1 44490: contig of 44490 bp in length
* 44491 44590: contig of 68101 bp in length
* 44591 112691: gap of 100 bp
* 112692 112791: gap of 100 bp
* 112792 149514: contig of 36723 bp in length
* 149515 149614: gap of 100 bp
* 149615 206355: contig of 56741 bp in length
* 206356 206456: gap of 100 bp
* 206456 219687: contig of 13232 bp in length.
FEATURES
    source
        1. 219687
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /db_xref="taxon:10090"
            /clone="RP23-75H23"
            /clone_id="RP23-75H23 Female Mouse BAC"
            1. 44490
            /note="assembly_fragment"
            clone_end:SP6
            vector_side:left"
        misc_feature
            44591..112691
            /note="assembly_fragment"
        misc_feature
            112792..149514
            /note="assembly_fragment"
        misc_feature
            149615..206355
            /note="assembly_fragment"
        misc_feature
            206456..219687
            /note="assembly_fragment"
            clone_end:T7
            vector_side:right"
BASE COUNT  69914 a 39620 c 40411 g 69342 t 400 others
ORIGIN
Query Match      0.9%; Score 21; DB 2; Length 219687;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1208 TTACCATCTGAACGTGTTAA 1228
Db 190487 TTACCATCTGAACGTGTTAA 190507

RESULT 97
AC098602 223266 bp DNA linear HTG 13-MAY-2003
DEFINITION Rattus norvegicus clone CH230-138H1, WORKING DRAFT SEQUENCE, 2
unordered pieces.
AC098602
AC098602.7 GI:30581472
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 223266)
Murny/D.Marie, Metzker,M.Lee, Abiraman,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalto,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,J., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,

```

```

Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Diaper,H., Dugan-Rocha,S., Dunn,A., Dubin,K., Duval,B., Eaves,K.,
Egan,A., Escoto,M., Eugene,C., Evans,C.A., Fall,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Georgievski,E., Geer,K., Gill,R., Garcia,M., Guerra,W., Guevara,W.,
Gunnarsson,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
Hollins,B., Howell,S., Huliy,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,U., Jacob,U., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Kapachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovat,C.,
Kowals,C., Kraft,C.L., Lebow,H., Levay,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorenshwa,L., Loulsegod,H., Lozada,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwankweme,O., Okwuon,G., Olarunpasegon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Plankoch,C.,
Plopper,F., Polindexter,A., Popovic,D., Primus,E., Pu,L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rivas,C., Rodkey,T., Roja,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smaiz,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sotelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Swalek,A., Taber,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,Y., Villaseña,D., Waldron,L., Walker,B., Wang,J.,
Wang,O., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wiczysk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 223266)
Worley,K.C.
Direct Submission
Submitted (26-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 223266)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
COMMENT
JOURNAL
AUTHORS
REFERENCE
JOURNAL
TITLE

```

The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GIKO

```

Center clone name: CH230-138H1
----- Summary Statistics -----
Assembly program: Atlas 3.0;
Consensus quality: 207235 bases at least Q40
Consensus quality: 209483 bases at least Q30
Consensus quality: 211150 bases at least Q20
Estimated insert size: 228805; sum-of-contigs estimation
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank\_draft\_data.html)
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 221213: contig of 221213 bp in length
* 221214 221313: gap of unknown length
* 221314 223266: contig of 1953 bp in length.
*
Location/Qualifiers
1..223266
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-138H1"
/complement(21319..219964)
/note="clone_boundary
clone_end:77
site:EcoRI
end_sequence:FMAMW377UB"
BASE COUNT 65373 a 44898 c 44784 g 57395 t 10816 others
ORIGIN

Query Match 0.9% Score 21; DB 2; Length 223266;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 2127 CAGTAGAGGAACACCTGCTG 2147
|||||
Db 10826 CAGTAGAGGAACACCTGCTG 10846

RESULT 98
AC106341 226427 bp DNA linear HTG 13-MAY-2003
AC106341 Rattus norvegicus clone CH230-154023,*** SEQUENCING IN PROGRESS
LOCUS
DEFINITION
AC106341
ACCESSION
AC106341
VERSION
AC106341.5 GI:30581326
KEYWORDS
HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 226427)
Muzny,D.Marie, Metzker,M.Lee, Adrianzon,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Angiulo,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,A., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswalto,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Butrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Cesari,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,D.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Dayila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denison,S., Deremo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flaigg,N., Forbes,L., Foster,M., Foster,P.,

```

Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
Gehrgeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, M., Guevara, W.,
Gunnarsson, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
Henderson, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M.,
Hollins, B., Howells, S., Huliy, S., Hume, J., Idlebird, D., Jackson, A.,
Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A.,
Karpachy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovac, C.,
Kowis, C., Kraft, C.U., Ledow, H., Levan, J., Lewits, L., Li, Z., Liu, J.,
Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
Lorenshwartz, L., Louisedge, H., Lozada, R. J., Lu, X., Ma, J.,
Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E.,
Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L.,
Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
Nwokedi, O., Okwuonu, G., Olarinmoye, A., Pal, S., Parks, K.,
Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankuch, C.,
Plopper, F., Poldexter, A., Popovic, D., Primus, E., Pu, L.,
Puzo, M., Quito, J., Rachin, E., Reeves, K., Regier, M.A., Reigh, R.,
Rell, B., Rell, J., Rell, J., Reuter, M., Richards, S., Riggs, F.,
Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.,
Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,
Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smaiz, D.,
Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J.,
Steinle, M., Strong, R., Sutton, A., Swalek, A., Taber, P., Taylor, C.,
Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K.,
Valas, R., Vera, V., Villaseca, D., Waldron, L., Walker, B., Wang, J.,
Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, F.,
Williams, G., Wilson, R., Wlezyk, R., Wooden, H., Worley, K.,
Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V.,
Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von
Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O.,
Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 226427)
Worley, K.C.

Direct Submission
Submitted (12-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 226427)

Rat Genome Sequencing Consortium.
Direct Submission
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

On May 13, 2003 this sequence version replaced gi:25007771.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GKSI
Center clone name: CH230-154023
----- Summary Statistics
Assembly program: Atlas 3.0,
Consensus quality: 211153 bases at least Q40

Consensus quality: 213655 bases at least Q30
 Consensus quality: 215176 bases at least Q20
 Estimated insert size: 223975; sum-of-coverage estimation
 Quality coverage: 6x in Q20 bases; sum-of-coverage estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 225286: contig of 225286 bp in length
 * 225287 225386: gap of unknown length
 * 225387 226427: contig of 1041 bp in length.
 Location/Qualifiers
 1..226427
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-154023"
 1..1438
 misc_feature
 /note="wgs contig"
 223249..225286
 /note="wgs_contig"
 BASE COUNT 59951 a 48875 c 48593 g 59078 t 9930 others
 ORIGIN

Query Match 0.9%; Score 21; DB 2; Length 226427;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2127 CAGTAGAGGAAACCTGCTG 2147
 |||||
 Db 147213 CAGTAGAGGAAACCTGCTG 147233

RESULT 99
 AC112436 236764 bp DNA linear HTG 15-NOV-2002
 LOCUS Rattus norvegicus clone CH230-153L4, *** SEQUENCING IN PROGRESS
 DEFINITION *** 4 unordered pieces.
 ACCESSION AC112436 GI:25006795
 VERSION HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
 KEYWORDS Rattus norvegicus (Norway rat)
 SOURCE Rattus norvegicus
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 236764)
 Muray,D,Marie, Metzker,M,lee, Abramson,S., Adams,C., Alder,J.,
 Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
 Ayala-Bechechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
 Bialow,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
 Cardenas,V., Carter,K., Cavazos,I., Caesar,H., Center,A.,
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., P.Souza,L.,
 Davis,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
 Diaper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
 Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
 Fraser,C.M., Gabisi,A., Gante,R., Garcia,A., Garner,T., Garza,M.,
 Gerstein,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
 Gunaratne,P., Haaland,W., Hamli,C., Hamilton,C., Hamilton,K.,
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hughes,M.,

Hollins,B., Howells,S., Hulv,S., Hume,J., Idlebird,D., Jackson,A.,
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Joliver,A.,
 Karpachy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Koyar,C.,
 Kovis,C., Kraft,C.U., Lebow,H., Levan,Z., Lewis,L., Li,Z., Liu,J.,
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
 Lorenz-Hewitt,L., Louisedge,H., Lozano,R.J., Lu,X., Ma,J.,
 Maheshwari,M., Mahdiantine,M., Mahmood,M., Malloy,K., Mangum,A.,
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
 Mawhney,S., McLeod,M.P., McNeill,T.Z., Meener,E.,
 Mlosoavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
 Morgan,M., Morris,K., Morris,S., Mundasa,M., Murphy,M., Nair,L.,
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
 Maokelameh,O., Okunolu,G., Olarunsoo,A., Pal,S., Parks,K.,
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
 Plopper,F., Poldexter,A., Popovic,D., Primus,E., Pu,L.,
 Puzo,M., Quiroz,J., Rachin,E., Reeves,K., Regier,M.A., Reigh,R.,
 Kelly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
 Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
 Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajic,D.,
 Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
 Steimle,M., Strong,R., Sutton,A., Swack,A., Tabor,P., Taylor,C.,
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Uemami,K.,
 Valas,R., Vera,V., Villaseana,D., Waldron,L., Walker,B., Wang,J.,
 Wang,O., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
 Williams,G., Willson,R., Wleczky,R., Wooden,H., Worley,K.,
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
 Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
 Weinstein,G. and Gibbs,R.A.

Direct Submission
 Unpublished
 2 (bases 1 to 236764)
 Worley,K.C.
 Direct Submission
 Submitted (21-FEB-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 236764)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Nov 15, 2002 this sequence version replaced gi:22856698.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/atlas/>). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu

----- Project Information
 Center project name: GRXO
 Center clone name: CH230-153L4
 Center Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 209856 bases at least Q40
 Consensus quality: 212207 bases at least Q30
 Consensus quality: 213866 bases at least Q20
 Estimated insert size: 213107; sum-of-coverage estimation
 Quality coverage: 6x in Q20 bases; sum-of-coverage estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 4887: contig of 4887 bp in length
 * 4888 4887: gap of unknown length
 * 233806 233806: contig of 228819 bp in length
 * 233807 233806: gap of unknown length
 * 233907 235015: contig of 1109 bp in length
 * 235016 235115: gap of unknown length
 * 235116 236764: contig of 1649 bp in length.
 Location/Qualifiers
 1. 236764
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-15314"
 1. 1012
 /note="wgs_contig"
 misc_feature 4988..6368
 /note="wgs_contig"
 misc_feature 231874..233805
 /note="wgs_contig"
 BASE COUNT 59222 a 48477 c 48404 g 58596 t 22065 others
 ORIGIN

Query Match 0.9% Score 21; DB 2; Length 236764;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2127 CAGTAGAGGAAACACCTGCTG 2147
 |||||
 155453 CAGTAGAGGAAACACCTGCTG 155473

RESULT 100
 AC130777/C
 LOCUS
 DEFINITION 4 unoriented pieces.
 AC130777
 AC130777.2 GI:22267876
 HTG; HTGS PHASE1; HTGS DRAFT; HTGS ENRICHED.
 Rattus norvegicus (Norway rat)
 SOURCE
 ORGANISM
 Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 240929)
 Muzny D.Marie, Metzger M.Lee, Abramson S., Adams C., Alder J.,
 Allen C., Allen H., Alsbrooks S., Amin A., Angiano D.,
 Anyalelechi V., Ayagi A., Ayodeji M., Beica E., Baden H.,
 Baldwin D., Bandanaike D., Barber M., Barnstead M., Benahmed F.,
 Bismalo K., Blair J., Blankenburg K., Blyth P., Brown M.,
 Bryant N., Buhay C., Burch P., Burrell K., Calderon E.,
 Cardenas V., Carter K., Cavazos I., Ceasar H., Center A.,
 Chacko J., Chaver D., Chen G., Chen R., Chen Y., Chen Z., Chu J.,
 Cleland C., Cockrell R., Cox C., Coyle M., Cree A., D'Souza L.,
 Davila M.L., Davis C., Davy-Carroll L., De Anda C., Dederich D.,
 Delgado O., Denson S., Deramo C., Ding Y., Dinh H., Divya K.,
 Draper H., Dugan-Rocha S., Dunn A., Durbin K., Duval B., Eaves K.,
 Egan A., Escotto M., Eugene C., Evans C.A., Falls T., Fan G.,
 Fernandez S., Finley M., Flagg N., Forbes L., Foster M., Foster P.,
 Frazer C.M., Gabisi A., Ganta R., Garcia A., Garner T., Garza M.,
 Gebregorjsgis E., Geer K., Gill R., Grady M., Guerra W., Guevara M.,
 Gunaratne P., Haaland W., Hamil C., Hamilton C., Hamilton K.,
 Harvey Y., Havlak P., Hawes A., Henderson N., Hernandez J.,

Hernandez R., Hines S., Hladun S.L., Hodgson A., Hognes M.,
 Hollins B., Howell S., Huik S., Hune D., Idubird D., Jackson A.,
 Jackson L., Jacob L., Jiang H., Johnson B., Johnson R., Jolyet A.,
 Kapathy S., Kelly S., Kelly S., Khan Z., King L., Kovar C.,
 Kowis C., Kraft C.L., Lebow H., Levan J., Lewis L., Li Z., Liu J.,
 Liu J., Liu W., Liu Y., London P., Longacre S., Lopez J.,
 Lorenshewa L., Louised H., Lozado R.D., Lu X., Ma J.,
 Maheshwari M., Mahidartene M., Mahmoud M., Malloy K., Mangum A.,
 Mangum B., Mapa P., Martin K., Martin R., Martinez E.,
 Mathiney S., McLeod M.P., McNeill T.Z., Meenan B.,
 Miosavljic A., Miner G., Minja E., Montemayor U., Moore S.,
 Morgan M., Morris K., Morris S., Mundasa M., Murphy M., Nat L.,
 Nankervis C., Neal D., Newton N., Nguyen N., Norris S.,
 Nwackeleme O., Okwong G., Olampunagoon A., Pal S., Parks K.,
 Pasternak S., Paul H., Perez A., Perez L., Pfankuch C.,
 Plopper F., Poindester A., Popovic D., Primus E., Pu J.,
 Puazo M., Quiroz J., Rachlin E., Reeves K., Reiter M.A., Reigh R.,
 Reilly B., Reilly M., Ren Y., Reuter M., Richards S., Riggs F.,
 Rivers C., Rodkey T., Rojas A., Rose M., Rose R., Ruiz S.J.,
 Sanders M., Savery G., Scherer S., Scott G., Shatman S., Shen H.,
 Shetty J., Shwartsbeyn A., Sison I., Sitter C.D., Smajls D.,
 Sneed J., Sodergren E., Song X.-Z., Sorelle R., Sosa J.,
 Steimle M., Strong R., Sutton A., Svatek A., Tabot P., Taylor C.,
 Taylor T., Thomas N., Thomas S., Tingey A., Trejos Z., Usmani K.,
 Valas R., Vera V., Villasana D., Waldron L., Walker B., Wang J.,
 Wang O., Wang S., Warren J., Warren R., Wei X., White F.,
 Williams G., Willson R., Wiczek R., Wooden H., Worley K.,
 Wright D., Wright R., Wu J., Yakub S., Yen J., Yoon L., Yoon V.,
 Yu F., Zhang J., Zhou J., Zhou X., Zhao S., Dunn D., von
 Niederhausern A., Weiss R., Smith D.R., Holt R.A., Smith H.O.,
 Weinstein G. and Gibbs R.A.

Direct Submission
 Unpublished
 2 (bases 1 to 240929)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (14-AUG-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 240929)
 Rat Genome Sequencing Consortium.
 Direct Submission
 Submitted (22-SEP-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Sep 22, 2002 this sequence version replaced gi:22218436.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). As a result, the
 sequence may extend beyond the ends of the clone and there may be
 contigs that consist entirely of whole genome shotgun sequence
 reads. Both end sequences and whole genome shotgun sequence only
 contigs will be indicated in the feature table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GNZE
 Center clone name: CH230-24M8
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990329
 Consensus quality: 21737 bases at least Q40
 Consensus quality: 220582 bases at least Q30
 Consensus quality: 222697 bases at least Q20
 Estimated insert size: 236179; sum-of-contigs estimation
 Quality coverage: 4x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 4 contigs. The true order of the pieces

* is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 10759: contig of 10759 bp in length
 * 10760 10859: gap of unknown length
 * 10860 120441: contig of 109582 bp in length
 * 120442 120541: gap of unknown length
 * 120542 239388: contig of 118847 bp in length
 * 239389 239488: gap of unknown length
 * 239489 240929: contig of 1441 bp in length.
 Location/Qualifiers

FEATURES
 source 1. 240929
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-24M8"
 1. 1105

misc_feature
 /note="wgs_end_extension
 clone_end:Sp6"
 6216. 7848

misc_feature
 /note="wgs_end_extension
 clone_end:Sp6"
 complement(7150. 8019)

misc_feature
 /note="clone_boundary
 clone_end:Sp6
 site:EcORI
 end_sequence:BH281801"
 49902. 49940

misc_feature
 /note="clone_boundary
 clone_end:T7
 site:EcORI
 end_sequence:BH281799"
 120542. 121801

misc_feature
 /note="wgs_end_extension
 clone_end:T7"

BASE COUNT 63052 a 47838 c 46729 g 66415 t 16895 others
 ORIGIN

Query Match 0.9%; Score 21; DB 2; Length 240929;
 Best Local Similarity 100.0%; Pred. No. 27;
 Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1919 ATTAATTGCTTGCTTGAT 1939
 |||||||

Db 45559 ATTAATTGCTTGCTTGAT 45539

Search completed: November 14, 2003, 06:25:58
 Job time : 8795 secs